

11/07/00

JC685 U.S. PTO

BERESKIN & PARR

BOX 401, 40 KING STREET WEST, TORONTO, CANADA M5H 3Y2
PHONE (416) 364-7311 • FAX (416) 361-1398 • WWW.BERESKINPARR.COM

PTO/SB/05 (08-00)

Please type a plus sign (+) inside this box → ☒

Approved for use through 10/31/2002. OMB 0851-0032

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY
PATENT APPLICATION
TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	9369-161
First Inventor	Maurice Meloney et al.
Title	OIL BODIES AND ASSOCIATED PROTEINS AS ..
Express Mail Label No.	

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

- ☐ Fee Transmittal Form (e.g., PTO/SB/17)
Submit an original and a duplicate for fee processing
- ☐ Applicant claims small entity status.
See 37 CFR 1.27.
- ☒ Specification [Total Pages 57]
(preferred arrangement set forth below)
 - Descriptive title of the invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to sequence listing, a table, or a computer program listing appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
- ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 24]
- Oath or Declaration [Total Pages]
 - ☐ Newly executed (original or copy)
 - ☐ Copy from a prior application (37 CFR 1.53 (d))
(for continuation/divisional with Box 17 completed)
 - ☐ **DELETION OF INVENTOR(S)**
Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.53(d)(2) and 1.53(b)
- ☐ Application Data Sheet. See 37 CFR 1.78

ADDRESS TO: Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

- ☐ CD-ROM or CD-R in duplicate, large table or Computer Program (Appendix)
- Nucleotide and/or Amino Acid Sequence Submission
(if applicable, all necessary)
 - ☐ Computer Readable Form (CRF)
 - Specification Sequence Listing on:
 - ☐ CD-ROM or CD-R (2 copies); or
 - ☐ paper
 - ☐ Statements verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

- ☐ Assignment Papers (cover sheet & document(s))
- ☐ 37 CFR 3.73(h) Statement ☐ Power of Attorney
(when there is an assignee)
- ☐ English Translation Document *(if applicable)*
- ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
- ☐ Preliminary Amendment
- ☐ Return Receipt Postcard (MPEP 503)
(should be specifically itemized)
- ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)
- ☐ Other:

17. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment, or in an Application Data Sheet under 37 CFR 1.75:

☐ Continuation ☐ Divisional ☒ Continuation-in-part (CIP)of prior application No. 09 / 318,275

Prior application information:

Examiner

Group / Art Unit: 1649

For CONTINUATION OR DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 5b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

18. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Labelor ☐ Correspondence address below

Name	Bereskin & Parr				
Address	Box 401				
	40 King Street West				
City	Toronto	State	Ontario	Zip Code	M5H 3Y2
Country	Canada	Telephone	(416) 364-7311	Fax	(416) 361-1398

Name (Print/Type)	Micheline Gravelle	Registration No. (Attorney/Agent)	40,261
Signature	<i>M. Gravelle</i>	Date	Nov. 7, 2000

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

JC921 U.S. PTO
09/707167
11/07/00

B&P File No. 9369-161

DERESKIN & PARR

UNITED STATES

Title: Oil Bodies and Associated Proteins as Affinity Matrices
Inventors: Maurice Moloney, Joseph Boothe and Gijs van Rooijen

TITLE: Oil Bodies and Associated Proteins as Affinity Matrices

This application is a continuation in part application of United States application serial no. 09/319,275 filed August 27, 1999 which is a continuation in part of United States application serial no. 08/767,026 filed
5 December 16, 1996, both of which are incorporated herein by reference.

FIELD OF THE INVENTION

This invention relates to the use of oil bodies and their associated proteins as affinity matrices for the separation and purification of target molecules from samples.

BACKGROUND OF THE INVENTION

Within the general field of biotechnology, the ability to effectively separate and purify molecules from complex sources, such as living cells, blood serum, or fermentation broth, is of critical importance. Applications in industry and research laboratories (where, for example,
15 purified or partly purified proteins are used) are numerous and well documented in prior literature. See, for example, R. Meadon and G. Walsh in *Biotechnological Advances* 1994, 12: pp 635-646.

The majority of currently employed techniques for the separation of molecules capitalizes on the innate physical and chemical properties of the molecule of interest. Affinity-based purification technologies are unique in that they exploit the highly specific biological recognition between two molecular species which form an affinity pair. Binding of the two entities of the affinity pair occurs in almost all instances as a result of relatively weak chemical interactions, known as non-covalent
25 bonds. Some art-recognized and commonly used affinity pairs include antibodies and their binding antigenic substances, nucleic acid binding proteins and nucleic acids, lipid binding proteins and lipids, lectins and carbohydrates, streptavidin/biotin complexes, protein A/immunoglobulin G complexes, and receptors and their binding molecules.

30 In general, affinity-based purification processes require that one member of the affinity pair is immobilized on a solid substrate or matrix that is insoluble in the fluid in which the other member of the pair

09707467 440700

resides. The molecular species of the affinity pair bound to the matrix is generally referred to as the ligand, while the liquid soluble member is generally referred to as the target member. However, it is important to note that these definitions do not impose any restrictions in a strict chemical sense. The vast majority of current ligand immobilization techniques rely on physical or chemical approaches. Physical ligand immobilization involves adsorption or entrapment of the ligand to a suitable support, while the chemical mode of immobilization is characterized by the formation of strong crosslinks or covalent attachments between the ligand and the matrix. It is a requirement that immobilization is accomplished in such a fashion that the capacity of the members of the affinity pair to recognize each other is not adversely affected by the immobilization procedure.

It is a disadvantage of the currently available physical and chemical techniques for immobilizing ligands that production processes are frequently time consuming and expensive. This is mainly due to the fact that immobilization techniques require the separate production of matrix material and ligands, which in a subsequent step must be coupled. An alternative mode of immobilizing proteins is described in U.S. Patent No. 5,474,925 which documents a biological production system for the immobilization of enzymes in the fibre of cotton plants. This patent discloses what is believed to be the first biologically produced enzyme immobilization system and allows a one step production of matrix and ligand.

Subsequent to immobilization of the ligand on the matrix, a variety of affinity based purification techniques may be employed to accomplish selective binding between the affinity immobilized ligand and the target member. Affinity based purification techniques known in the prior art include perfusion affinity chromatography, affinity repulsion chromatography, hyperdiffusion affinity chromatography, affinity precipitation, membrane affinity partitioning, affinity cross-flow ultrafiltration and affinity precipitation. In the most widely used affinity based purification technique, affinity chromatography, a matrix containing a ligand is coated to, or packed on, the inside of a chromatographic column. A complex mixture containing the target member is then applied to the

- 3 -

chromatographic column. Ideally, only the target molecules that specifically recognize the ligand bind in a non-covalent fashion to the chromatographic column, while all other molecular species present in the sample pass through the column.

5 In affinity partitioning, two solutions of substantially different densities are employed. The complex solution containing the target member is mixed with a solution of a different density containing the affinity ligand. Subsequent to mixing, the solutions are left to settle in order to permit the formation of two separate phases. Molecules tend to partition
10 differentially between phases depending on their size, charge and specific interactions with the phase-forming solutions. Ligand-bound target protein selectively partitions to the phase containing the affinity ligand. For example, Coughlin and Bacilaski in *Biotechnology Progress*, 1990 6: 307-309 reported the use of the biotin containing organic solution isooctane to
15 transfer avidin from an aqueous solution to the isooctane solution. However, so far applications of affinity partitioning have been limited mainly due to the current lack of availability of suitable affinity matrix substances which can be employed in specific partitioning in two phase systems.

20 An important factor for the commercial development of biotechnology is the purification of bioproducts, which typically accounts for 50% or more of the total costs (Labrou, N. and Clonis, Y. D. in the *Journal of Biotechnology* 36: 95-119 (1994)). Many protein purification steps rely on column type separation procedures. In particular, large scale high-
25 separation techniques such as column chromatography or batch-type based protein purification techniques are costly. In addition, crude material is less suitable for either column chromatography or batch separations, as contaminants may foul up sedimented resins and plug columns. Thus, affinity matrices are often only employed in a later stage of purification
30 processes where substantial purity is critical, where the proteins are present in extremely dilute concentrations, or where high value proteins are required, for example in diagnostic and therapeutic proteins. These and other topics related to the use of affinity technology in biotechnological

- 4 -

processes have been reviewed by Labrou, N. and Clonis, Y. D. in the *Journal of Biotechnology* 36: 95-119 (1994).

There is a need in the art to develop novel and economical methods for separating and purifying biological products from complex mixtures. The present inventors have found that subcellular oil storage structures, known as oil bodies, and their associated proteins are useful in this regard.

SUMMARY OF THE INVENTION

The present invention relates to a novel versatile biological system for the production of affinity matrices. The present inventors have found that oil bodies and their associated proteins can be used as affinity matrices for the separation of a wide variety of target molecules such as proteins, carbohydrates, lipids, organic molecules, nucleic acids, metals, cells and cell fractions from a sample.

In accordance with the invention, there is provided a method for the separation of a target molecule from a sample comprising: 1) contacting (i) oil bodies that can associate, either directly or indirectly, with the target molecule with (ii) a sample containing the target molecule; and 2) separating the oil bodies associated with the target molecule from the sample. The oil bodies and the sample containing the target molecule are brought into contact in a manner sufficient to allow the oil bodies to associate with the target. Preferably, oil bodies are mixed with the target. If desired, the target molecule may be further separated from the oil bodies.

In one aspect, the target molecule has affinity for, or binds directly to, the oil bodies or oil body protein. Examples of such targets include antibodies or other proteins that bind to oil bodies.

In another aspect, a ligand molecule may be used to associate the target molecule with the oil bodies.

In one embodiment, the ligand has natural affinity for the oil bodies or oil body protein. In a specific embodiment, the ligand is an antibody that binds the oil body protein. Such an antibody can be used to separate targets having natural affinity for the ligand antibody such as anti-IgG antibodies or protein A. A bivalent antibody may also be prepared having binding specificities for both the oil body protein and the target. The

5

10

20

25

30

Other objects, features and advantages of the present invention will become apparent from the following detailed description and

attached drawings. It should be understood, however, that the detailed description and associated examples are given by way of illustration only, and various changes and modifications thereto falling within the scope of the invention will become apparent to those skilled in the art. In addition, reference is made herein to various publications, patents and patent applications which are hereby incorporated by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. The nucleotide and deduced amino acid sequence of the 18 KDa oleosin from *Arabidopsis thaliana* as shown in SEQ.ID.NO:1 and SEQ.ID.NO:2.

Figure 2. Sequence of an *Arabidopsis* oleosin-hirudin fusion. Indicated are a portion of the oleosin genomic sequence (from base 1-1620 as reported in van Rooijen *et al* 1992, *Plant Mol. Biol.* 18: 1177-1179), a spacer sequence (base 1621-1635, underlined) and the synthetic DNA sequence encoding the mature hirudin variant-2 isoform (base 1636-1833, italicized). This gene fusion is regulated by the 5' upstream region of the *Arabidopsis* oleosin (bases 1-861) and the nopaline synthase termination sequence (base 1855-2109). The sequence is also shown in SEQ.ID.NO:3 and SEQ.ID.NO:4.

Figure 3. Outline of the steps employed in the construction of pCCGOBHIRT, containing the entire oleosin-hirudin construct.

Figure 4. Schematic diagram illustrating the configuration of the oleosin-hirudin fusion protein on the oil body and the binding of thrombin.

Figure 5. NaCl elution profiles of thrombin from wild type and 4A4 oil body matrices transformed with a construct expressing an oleosin-hirudin fusion.

Figure 6. Purification of a horseradish peroxidase conjugated anti-IgG antibody using an anti-oleosin antibody as a ligand. Schematic diagram illustrating the configuration of the oleosin / anti-oleosin / anti-IgG sandwich complex bound to an oil body.

Figure 7. Illustrates specific binding of anti-IgG antibodies to wild type oil bodies complexed with primary anti-oleosin antibodies as a ligand (left) and binding of anti-IgG antibodies to oil bodies which were not

complexed with primary antibodies prior to binding with the secondary antibodies (right).

Figure 8. Sequence of an oleosin metallothionein fusion. Indicated are the coding sequence of a *B. napus* oleosin cDNA (bases 1092-1652, van Rooijen, 1993, *Ph.D. Thesis*, University of Calgary), a spacer sequence (bases 1653-1670, underlined) and the human metallothionein gene *mt-II* (bases 1671-1876, Varshney and Gedamu, 1984, *Gene*, 31: 135-145)). The gene fusion is regulated by an *Arabidopsis* oleosin promoter (bases 1-1072) and ubiquitin termination sequence (bases 1870-2361, *ubi3'*; Kawallerck *et al.*, 1993, *Plant Mol. Biol.* 21: 673-684). The sequence is also shown in SEQ.ID.NO:6 and SEQ.ID.NO:7.

Figure 9. Outline of the steps employed in the construction of pR100M3' containing the entire oleosin-metallothionein construct.

Figure 10. Schematic diagram illustrating the configuration of the oleosin-metallothionein fusion protein on the oil body and binding of cadmium ions.

Figure 11. Illustrates the binding (A) and elution (B) of cadmium to an oil body matrix from wildtype *B. carinata* seeds and *B. carinata* seeds transformed with a construct expressing oleosin metallothionein gene fusion. Shown is the percentage cadmium bound to the oil body fraction of an oil body fraction harvested from transgenic and untransformed seeds. Bars represent average values of 5 replicate experiments (binding) and 3 replicates (elution).

Figure 12. Illustrates the binding of protein A expressing *S. aureus* cells to oil bodies treated with varying amounts of anti-oleosin IgGs. Bars represent OD₆₀₀ readings obtained following the procedures as described in Example 5 and using varying amounts of IgGs (0 µl, 3 µl, 30 µl, 100 µl of added IgG).

Figure 13. Oligonucleotide primers used to amplify the sequence of the *S. aureus* protein A (The sequence is also shown in SEQ.ID.NO:8; The protein sequence is also shown in SEQ.ID.NO:9). Primer BK266, 5'C TCC ATG GAT CAA CGC AAT GGT TTA TC 3' (SEQ.ID.NO:10), a *NcoI* site (italicized) and a sequence identical to a portion of the protein A gene as contained within vector pR11221 (Pharmacia) (underlined) are

004076 4374260

indicated. Primer BK267, 5' GC AAG CTT CTA ATT TGT TAT CTG CAG
GTC 3' (SEQ.ID.NO:11), a *HindIII* site (italicized), a stop codon (bold) and a
sequence complementary to a portion of the protein A gene as contained
within pRIT2T (Pharmacia) (underlined) are indicated. The PCR product
5 was digested with *NcoI* and *HindIII* and ligated into pCGNOBPGUSA (Van
Rooijen and Moloney, 1995, *Plant Physiol.* 109: 1353-1361) from which the
NcoI-GUS-*HindIII* fragment had been removed.

Figure 14. Sequence of an *Arabidopsis* oleosin-protein A fusion
(The sequence is also shown in SEQ.ID.NO:12 and the protein sequence is
10 also shown in SEQ.ID.NO:13 and 14). Indicated are a portion of the oleosin
genomic sequence (from base 1 - 1626, as reported in van Rooijen *et al.*, 1992
Plant Mol. Biol. 18: 1177-1179), a spacer sequence encoding a thrombin
cleavage site (base 1627 - 1647, underlined) and the DNA sequence encoding
protein A (base 1648 - 2437, italicized). Expression is regulated by the
15 *Arabidopsis* 5' upstream region of the *Arabidopsis* oleosin (base 1 - 867) and
the nopaline synthase terminator region (base 2437 - 2700).

Figure 15. Schematic diagram illustrating the configuration of
the oleosin-protein A fusion protein on the oil body and binding of the
immunoglobulin.

20 Figure 16. A western blot illustrating the binding of HRP-
conjugated mouse anti-rabbit antibodies to oil body protein extracts
obtained from transgenic *B. napus* lines expressing oleosin-protein A fusion
proteins. Shown on a Western blot probed with an HRP- conjugated
antibody are oil body protein extracts from transgenic lines, opa 30 (lane 3),
25 opa 31 (lane 4), opa 34 (lane 5), opa 36 (lane 6), opa 47 (lane 7), opa 93 (lane
8), all expressing an oleosin-protein A fusion protein and a control
untransformed *B. napus* line (lane 9), as well as lysates of *E. coli* DH5 α
transformed with pRIT2T expressing protein A (lane 2) and untransformed
E. coli DH5 α (lane 1).

30 Figure 17. illustrates binding and elution of IgGs to oil bodies
isolated from wildtype *B. napus* (bn wt) and a transgenic *B. napus* line,
expressing an oleosin protein A fusions. Error bars represent the results
from 4 independent experiments.

Figures 18A-C. Nucleotide sequence of the phaseolin promoter-PRS-OBScFv-Prochymosin -phaseolin terminator sequence (SEQ ID No. 15). The phaseolin promoter corresponds to nucleotide 6-1554. The DNA sequence encoding the PRS-OBScFv-Prochymosin gene fusion corresponds to nt 1554-3467. The phaseolin terminator corresponds to nucleotide sequence 3474-4694. The deduced amino acid sequence of the PRS-OBScFv-Prochymosin fusion is also indicated. Met1 to Ala25 corresponds to the PRS signal sequence, Glu28 to Thre 142 corresponds to the Variable heavy antibody chain, Gly 143 to Ser156 corresponds to the flexible linker peptide separating Vh and Vl, Asp157-Leu271 corresponds to the Variable light chain and Ala273 to Ile 638 corresponds to the prochymosin peptide.

Figure 19. Western Blot of seed and oil body extracts of *Arabidopsis* plants transformed with pSB52168. A total of 12.5 µl of sample was loaded in each lane. Lane 1; Bovine chymosin, Lane 2; total extract, Lane 3; Sup1, Lane 4; OB0, Lane 5 ;OBHighS. See the text for a description of the samples. This Western blot was treated with polyclonal antibodies raised against bovine chymosin followed by an alkaline phosphatase linked secondary antibody and NBT/BCIP color reaction. Indicated are the OBScFvProchymosin protein fusion (OBScFvProchymosin), a band which could correspond to an aberrantly folded OBScFvProchymosin protein fusion, and as a result has a slightly slower mobility on the polyacrylamide gel and no detectable affinity to *Arabidopsis* oil bodies. Also indicated are the bands that correspond to the mature processed form of chymosin (chymosin) and proteolytic breakdown products of chymosin (partial chymosin).

DETAILED DESCRIPTION OF THE INVENTION

As hereinbefore mentioned, the present invention relates to a novel biological affinity matrix system that employs oil bodies and their associated proteins. The affinity matrix is suitable for the highly-efficient separation of specific targets, including proteins, carbohydrates, lipids, nucleic acids, cells and subcellular organelles, metals and ions, from a sample.

10 -

The present invention provides a method for the separation of a target molecule from a sample comprising: 1) contacting (i) oil bodies that can associate either directly or indirectly with the target molecule with (ii) a sample containing the target molecule; and 2) separating the oil bodies associated with the target molecule from the sample. The oil bodies and the sample containing the target molecule are brought into contact in a manner sufficient to allow the oil bodies to associate with the target. Preferably, the oil bodies are mixed with the target. Indirect association of the oil bodies with the target can be effected using a ligand molecule that can associate with both the oil bodies and the target molecule. The ligand therefore serves to bridge or join the oil bodies with the target molecule. If desired, the target molecule may be further separated from the oil bodies and the ligand, if present.

Each of the components of the affinity matrix are discussed in turn below.

Targets

The term "target" as used herein denotes a desired molecule that one wants to purify, isolate or separate from a sample such as a biological mixture. This technology is amenable for use with virtually any target for which a ligand can be obtained or any target that can directly associate with or bind to an oil body or oil body protein. Possible ligand/target pairs include but are not limited to: protein subunit/subunit associations, antibodies/antigens, receptor protein/signal molecules, nucleic acid binding proteins/nucleic acids; lectins/carbohydrates; lipid binding proteins/lipids; ion binding proteins/ions; and ligands to surface epitopes/cells or subcellular organelles. The target may be obtained from any natural source or may be synthesized chemically. If the target is a macromolecule such as a protein or nucleic acid it may also be produced in recombinant form using any suitable expression system such as bacteria, yeast, plant, insect, mammalian, etc.

Ligands

The term "ligand" used herein denotes a molecule that is capable of associating with both the target molecule and the oil bodies or oil body protein (discussed below). The term "associating with" as used herein

includes both covalent and non-covalent binding of the ligand to the oil bodies or the target molecule. For example, the ligand molecule may be covalently attached to the oil bodies (or oil body protein) and non-covalently associate with the target (and vice-versa), or the ligand may non-covalently associate with both the oil bodies and the target molecule. The ligand may be any molecule that can bridge the oil bodies or oil body protein and the target molecule and can include a protein, nucleic acid, carbohydrate or small organic molecule. The ligand may be comprised of two molecules, a first molecule that associates with the oil bodies and a second molecule that associates with the target, wherein the first molecule and the second molecule associate with each other.

The affinity ligand proteins used for this methodology may be derived from naturally-occurring, known ligand pairs such as those listed above. Alternatively, the ligand may be obtained by screening proteins extracted from cells or organisms, synthesized chemically or produced in libraries comprised of combinatorial peptide sequences, antibodies, or expressed DNA sequences.

In one embodiment, the ligand has natural affinity for the oil bodies or the oil body protein. For example, the ligand may be a protein such as an antibody, that has affinity for the oil body protein. The ligand may also be a molecule other than a protein which has natural affinity for the oil body or oil body protein. Such ligands, capable of binding to the oil bodies or oil body protein, may be associated either directly or indirectly with the target molecule. In a particular embodiment, the ligand is covalently attached to the target molecule by chemical or recombinant means. For example, the ligand may be an antibody that is prepared as a recombinant fusion protein with the target. The ligand may also be associated with a second molecule that can bind the target molecule. For example, the ligand molecule may be an antibody conjugated to avidin and can be used to purify biotin from a sample.

In another embodiment, the ligand is covalently linked to the oil bodies or oil body protein by chemical or recombinant means. Chemical means for preparing fusions or conjugates are known in the art and can be used to prepare a ligand-oil body protein fusion. The method used to

- 12 -

conjugate the ligand and oil body must be capable of joining the ligand with the oil body protein without interfering with the ability of the ligand to bind to the target molecule. In one example, the ligand may be a small organic molecule such as biotin that is covalently attached to the oil bodies.

5 Biotinylated oil bodies can be used to separate avidin from a sample. The present invention also includes modified oil bodies such as biotinylated oil bodies for use as an affinity matrix. Accordingly, the present invention includes a composition comprising oil bodies attached to a molecule, such as a ligand or a target molecule.

10 In a preferred embodiment, the ligand is a protein and can be conjugated to the oil body protein using techniques well known in the art. There are several hundred crosslinkers available that can conjugate two proteins. (See for example "Chemistry of Protein Conjugation and Crosslinking", 1991, Shans Wong, CRC Press, Ann Arbor). The crosslinker
15 is generally chosen based on the reactive functional groups available or inserted on the ligand. In addition, if there are no reactive groups a photoactivatable crosslinker can be used. In certain instances, it may be desirable to include a spacer between the ligand and the oil-body protein. Crosslinking agents known to the art include the homobifunctional agents:
20 glutaraldehyde, dimethyladipimide and Bis(diazobenzidine) and the heterobifunctional agents: *m*-Maleimidobenzoyl-*N*-Hydroxysuccinimide and Sulfo-*m*-Maleimidobenzoyl-*N*-Hydroxysuccinimide.

A ligand protein-oil body protein fusion may also be prepared using recombinant DNA techniques. In such a case a nucleic acid sequence
25 encoding the ligand is fused to a nucleic acid sequence encoding the oil body protein, resulting in a chimeric nucleic acid molecule that expresses a ligand-oil body protein fusion protein (discussed in greater detail below). In order to prepare a recombinant fusion protein, the sequence of the nucleic acid encoding the ligand must be known or be obtainable. By obtainable it is
30 meant that a nucleic acid sequence sufficient to encode the protein ligand may be deduced from the known amino acid sequence. It is not necessary that the entire gene sequence of the ligand be used provided that a subsequence encoding the binding domain of the protein ligand is known.

004037 2920460

Therefore, the ligand can include the complete sequence of, or the binding domain from, the specific ligand protein in question.

If the nucleic acid sequence of the desired ligand is known, the gene may be synthesized chemically using an oligonucleotide synthesizer.

5 Alternatively, the clone carrying the ligand gene may be obtained from either cDNA or genomic libraries containing the gene by probing with a labelled complementary nucleic acid sequence. The gene may also be specifically amplified from the library using gene-specific oligonucleotide primers and the PCR. If the nucleic acid sequence of the desired ligand is
10 not known, then a partial amino acid sequence may be obtained through N-terminal sequencing of the protein (Matsudaira 1987; *J. Biol. Chem.* 262: 10035-10038). Labelled probes may be synthesized based upon the DNA sequences deduced from this amino acid sequence and used to screen cDNA or genomic libraries as described above. The clone carrying the gene may
15 also be identified from a cDNA expression library by probing either with antibodies raised against the protein ligand, or with the target protein

Ligands may also be uncovered by probing mixtures of proteins with the target. The target can be immobilized on a support matrix and used to screen proteins extracted from cells and tissues or synthesized
20 chemically. Following binding between the ligand protein and the immobilized target, the matrix is separated from the solution and washed. The protein ligand is subsequently eluted from the matrix and the sequence determined as described above. Alternatively, recombinant protein libraries produced by phage display, such as those comprised of combinatorial
25 peptide sequences (Smith, 1985; *Science* 228: 1315-1317) or antibody repertoires (Griffiths *et al.*, 1994, *EMBO J.* 13: 3245-3260, Nissim *et al.*, 1994, *EMBO J.* 13: 692-698) can be screened with the immobilized target. In this case, binding between the protein ligand and the target would enable separation and recovery of the phage expressing the ligand from the large,
30 complex population of phage encoding non-binding proteins. A two-hybrid system such as that in yeast (Fields and Sternglanz, 1994; *Trends Genet.* 10: 286-292) might also be used to identify a ligand from an expressed cDNA library. Here, a gene fusion is constructed between the sequence encoding the target protein and that of a DNA binding protein. Cells containing this

construct are transformed with constructs from a cDNA library where the sequences have been fused to that of a transcriptional activator. Binding between ligands derived from the cDNA library with the target protein allows transcription of a reporter gene to occur. Clones expressing the
5 ligand are then recovered.

To specifically uncover a ligand to oil bodies, a complete or partial oleosin protein may be used as target in any of the above methods. Alternatively, it may be possible to employ intact oil bodies for screening protein extracts, synthetic peptides or phage display libraries. In this case,
10 the oil body would serve both as target and immobilization matrix. Using this approach, a wider variety of ligands may be uncovered; that exhibit affinity not only to oleosins, but to other epitopes present on oil bodies.

Oil bodies and Oil Body Proteins

Oil bodies are small, spherical, subcellular organelles
15 encapsulating stored triacylglycerides, an energy reserve used by many plants. Although they are found in most plants and in different tissues, they are particularly abundant in the seeds of oilseeds where they range in size from under one micron to a few microns in diameter. Oil bodies are comprised of the triacylglycerides surrounded by a half-unit membrane of
20 phospholipids and embedded with a unique type of protein known as an oil body protein. The term "oil body" or "oil bodies" as used herein includes any or all of the triacylglyceride, phospholipid or protein components present in the complete structure. The term "oil body protein" as used herein means a protein that is naturally present in an oil body. In plants, the
25 predominant oil body proteins are termed "oleosins". Oleosins have been cloned and sequenced from many plant sources including corn, rapeseed, carrot and cotton. The oleosin protein appears to be comprised of three domains; the two ends of the protein, N- and C-termini, are largely hydrophilic and reside on the surface of the oil body exposed to the cytosol
30 while the highly hydrophobic central core of the oleosin is firmly anchored within the membrane and triacylglyceride. Oleosins from different species represent a small family of proteins showing considerable amino acid sequence conservation, particularly in the central region of protein. Within an individual species, a small number of different isoforms may exist.

Oil bodies from individual species exhibit a roughly uniform size and density which is dependent in part upon the precise protein/phospholipid/triacylglyceride composition. As a result, they may be simply and rapidly separated from liquids of different densities in which they are suspended. For example, in aqueous media where the density is greater than that of the oil bodies, they will float under the influence of gravity or applied centrifugal force. In 95% ethanol where the density is less than that of the oil bodies, they will sediment under the same conditions. Oil bodies may also be separated from liquids and other solids present in solutions or suspensions by methods that fractionate on the basis of size. For example, the oil bodies from *B. napus* are minimal, approximately 0.5µm in diameter, and thus may be separated from smaller components using a membrane filter with a pore size less than this diameter.

The oil bodies of the subject invention are preferably obtained from a seed plant and more preferably from the group of plant species comprising: thale cress (*Arabidopsis thaliana*), rapeseed (*Brassica* spp.), soybean (*Glycine max*), sunflower (*Helianthus annuus*), oil palm (*Elaeis guineensis*), cottonseed (*Gossypium* spp.), groundnut (*Arachis hypogaea*), coconut (*Cocos nucifera*), castor (*Ricinus communis*), safflower (*Carthamus tinctorius*), mustard (*Brassica* spp. and *Sinapis alba*), coriander (*Coriandrum sativum*) linseed/flax (*Linum usitatissimum*), and maize (*Zea mays*). Plants are grown and allowed to set seed using agricultural cultivation practises well known to a person skilled in the art. After harvesting the seed and removal of foreign material such as stones or seed hulls, for by example sieving, seeds are preferably dried and subsequently processed by mechanical pressing, grinding or crushing. The oil body fraction may be obtained from the crushed seed fraction by capitalization on separation techniques which exploit differences in density between the oil body fraction and the aqueous fraction, such as centrifugation, or using size exclusion-based separation techniques, such as membrane filtration, or a combination of both of these. Typically, seeds are thoroughly ground in five volumes of a cold aqueous buffer. A wide variety of buffer compositions may be employed, provided that they do not contain high concentrations of strong organic solvents such as acetone or diethyl ether, as these solvents may disrupt the oil bodies. The

solution density of the grinding buffer may be increased with the addition of 0.4-0.6 M sucrose, in order to facilitate washing as described below. The grinding buffer will also typically contain 0.5 M NaCl to help remove soluble proteins that are not integrally bound to the oil body surface.

5 Following grinding, the homogenate is centrifuged resulting in a pellet of particulate and insoluble matter, an aqueous phase containing soluble components of the seed, and a surface layer comprised of oil bodies with their associated proteins. The oil body layer is skimmed from the surface and thoroughly resuspended in one volume of fresh grinding
10 buffer. It is important that aggregates of oil bodies are dissociated as thoroughly as possible in order to ensure efficient removal of contaminants in the subsequent washing steps. The resuspended oil body preparation is layered under a floatation solution of lower density (e.g. water, aqueous buffer) and centrifuged, again, separating oil body and aqueous phases.
15 This washing procedure is typically repeated at least three times, after which the oil bodies are deemed to be sufficiently free of contaminating soluble proteins as determined by gel electrophoresis. It is not necessary to remove all of the aqueous phase and to the final preparation water or 50 mM Tris-HCl pH 7.5 may be added and if so desired the pH may be lowered to pH 2
20 or raised to pH 10. Protocols for isolating oil bodies from oil seeds are available in Murphy, D. J. and Cummins I., 1989, *Phytochemistry*, 28: 2063-2069; and in: Jacks, T. J. et al., 1990, *JAOCS*, 67: 353-361. A preferred protocol is detailed in example 1 of the present specification.

Oil bodies other than those derived from plants may also be
25 used in the present invention. A system functionally equivalent to plant oil bodies and oleosins has been described in bacteria (Pieper-Fürst et al., 1994, *J. Bacteriol.* 176: 1328), algac (Rossler, P.G., 1988, *J. Physiol. (London)*, 24: 394-400) and fungi (Ting, J. T. et al., 1997, *J. Biol Chem.* 272: 3699-3706). Oil bodies from these organisms, as well as those that may be discovered in other
30 living cells by a person skilled in the art, may also be employed according to the subject invention.

Affinity Matrices

As hereinbefore mentioned, the present invention provides a novel affinity matrix system for the purification of a target molecule from a

sample. In one embodiment, the affinity matrix comprises oil bodies that can bind a target molecule in a sample. In such an embodiment, the target molecule may be an antibody that can bind an oil body protein. In another embodiment, the affinity matrix comprises oil bodies or oil body proteins and a ligand that is associated with the oil bodies or oil body proteins and has affinity for a target molecule. In such an embodiment, the ligand may be non-covalently or covalently attached to the oil bodies or oil body protein (as described above). In another embodiment, the affinity matrix comprises oil bodies or oil body proteins and a ligand that is covalently or non-covalently attached to the target.

It is an advantage of the present invention that target substances can be purified or removed from samples through non-covalent association with oil bodies followed by oil body separation. A number of different oil body-ligand configurations are possible. Targets with inherent affinity for a specific ligand proteins such as hirudin to thrombin or heavy metals to metallothionein, may be purified or separated with oil bodies containing that ligand fused to an oleosin. Alternatively, a protein target may also be purified or separated with an oil body affinity matrix by fusing the target to an oil body-specific ligand (such as an antibody that binds the oil bodies or oil body proteins) or to a ligand complimentary to that fused to an oleosin. If desired, a protease recognition site or chemical cleavage site may be engineered between the ligand and the target protein to enable proteolytic removal of the ligand from the target protein in the course of purification. A multivalent ligand may also be constructed, such as a bivalent single-chain antibody, in which one domain of the ligand has an affinity for an oil body and the other domain(s) exhibits affinity for the target. In this case, neither the oil body nor the target molecule need to be covalently fused to a ligand. Also, concatamers of ligands may be used to increase the affinity of a matrix for a target, or the sequence of a ligand may be mutated to modulate the affinity for a target when such conditions are desirable. Further, mixtures of different ligands may be fused to recover/remove different types of targets simultaneously. Fusions between different ligands may also be constructed to form bridges between different types of targets or between targets and the oil body affinity

matrix. Binding to the affinity matrix may also be achieved by forming bridges between ligand or ligand and target sequences, such as Zn^{++} ions bridging between polyhistidine sequences.

There are several advantages associated with the use of oil body affinity matrices that make them attractive as purification tools. The flexibility in design that is possible through the different configurations described above, enables a matrix to be constructed to best meet the requirements for a specific target. Also, production of the matrix as part of a natural biological process in seeds is extremely cost-effective, since purification and immobilization of the ligand are not necessary. In the case of oleosin-ligand fusions, the ligand is immobilized on the oil body as a result of oleosin targeting within the cell, while oil body-specific ligands will naturally associate with the matrix while present in complex mixtures. Natural immobilization of the ligand on the matrix may also be advantageous in that it eliminates the requirement for chemical cross-linking that may compromise the affinity of the ligand for the target. Finally, oil body affinity matrices offer a unique and attractive purification option particularly for large scale operations. The ability to separate the matrix through floatation as a loose suspension enables it to be employed with crude material containing what might otherwise be prohibitive amounts of particulate contaminants. The presence of these contaminants will often foul and block conventional solid matrices applied in columns or batch suspensions limiting their use at early stages in the purification process.

As mentioned previously, in one embodiment of the invention, ligand protein sequences are genetically fused to the oil body protein. In order to prepare such genetic fusions, a chimeric nucleic acid sequence is prepared that encodes an oil body protein-ligand fusion protein and comprises (a) a nucleic acid sequence encoding a sufficient portion of an oil body protein to provide targeting of the fusion protein to the oil bodies and (b) a nucleic acid sequence encoding a sufficient portion of the ligand protein to provide binding of the target. The inventors have determined that, in general, the N-terminus and the hydrophobic core of an oil body protein are sufficient to provide targeting of the fusion protein to the oil

The coding region of the construct will typically be comprised of sequences encoding a ligand fused in frame to an oleosin and ending with a translational termination codon. The sequence for the oleosin may be

5
10

15

20

25

30

A. tumefaciens or *A. rhizogenes* to allow for transfer of the transcription construct to the plant cells. Following transformation using *Agrobacterium* the plant cells are dispersed in an appropriate medium for selection, subsequently callus, shoots and eventually plantlets are recovered. The *Agrobacterium* host will harbour a plasmid comprising the *vir* genes necessary for transfer of the T-DNA to the plant cells. For injection and electroporation, (see below) disarmed Ti-plasmids (lacking the tumour genes, particularly the T-DNA region) may be introduced into the plant cell.

The use of non-*Agrobacterium* techniques permits the use of the constructs described herein to obtain transformation and expression in a wide variety of monocotyledonous and dicotyledonous plants and other organisms. These techniques are especially useful for species that are intractable in an *Agrobacterium* transformation system. Other techniques for gene transfer include biolistics (Sanford, 1988, *Trends in Biotech.*, 6: 299-302), electroporation (Fromm *et al.*, 1985, *Proc. Natl. Acad. Sci. USA*, 82: 5824-5828; Riggs and Bates, 1986, *Proc. Natl. Acad. Sci. USA* 83: 5602-5606) or PEG-mediated DNA uptake (Potrykus *et al.*, 1985, *Mol. Gen. Genet.*, 199: 169-177).

In a specific application, such as to *B. napus*, the host cells targeted to receive recombinant DNA constructs typically will be derived from cotyledonary petioles as described by Moloney *et al.*, (1989, *Plant Cell Rep.*, 8: 238-242). Other examples using commercial oil seeds include cotyledon transformation in soybean explants (Hinchey *et al.*, 1988, *Bio/Technology*, 6: 915-922) and stem transformation of cotton (Umbeck *et al.*, 1981, *Bio/Technology*, 5: 263-266).

Following transformation, the cells, for example as leaf discs, are grown in selective medium. Once shoots begin to emerge, they are excised and placed onto rooting medium. After sufficient roots have formed, the plants are transferred to soil. Putative transformed plants are then tested for presence of a marker. Southern blotting is performed on genomic DNA using an appropriate probe, for example an *A. thaliana* oleosin gene, to show that integration of the desired sequences into the host cell genome has occurred.

The expression cassette will normally be joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a

herbicide, e.g. phosphinothricin or glyphosate, or more particularly an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, or the like. The particular marker employed will be one which will allow for selection of transformed cells compared with cells
5 lacking the introduced recombinant DNA.

The fusion peptide in the expression cassette constructed as described above, expresses at least preferentially in developing seeds. Accordingly, transformed plants grown in accordance with conventional ways, are allowed to set seed. See, for example, McCormick et al. (1986,
10 *Plant Cell Reports*, 5: 81-84). Northern blotting can be carried out using an appropriate gene probe with RNA isolated from tissue in which transcription is expected to occur, such as a seed embryo. The size of the transcripts can then be compared with the predicted size for the fusion protein transcript.

Oil body proteins are then isolated from the seed and analyses performed to determine that the fusion peptide has been expressed. Analyses can be for example by SDS-PAGE. The fusion peptide can be detected using an antibody to the oleosin portion of the fusion peptide. The size of the fusion peptide obtained can then be compared with predicted size
20 of the fusion protein.

Two or more generations of transgenic plants may be grown and either crossed or selfed to allow identification of plants and strains with desired phenotypic characteristics including production of recombinant proteins. It may be desirable to ensure homozygosity of the plants, strains
25 or lines producing recombinant proteins to assure continued inheritance of the recombinant trait. Methods of selecting homozygous plants are well known to those skilled in the art of plant breeding and include recurrent selfing and selection and anther and microspore culture. Homozygous plants may also be obtained by transformation of haploid cells or tissues
30 followed by regeneration of haploid plantlets subsequently converted to diploid plants by any number of known means, (e.g.: treatment with colchicine or other microtubule disrupting agents).

Method of Separating Target Molecules Using the Affinity Matrices

As hereinbefore mentioned, the present invention relates to a method of separating a target molecule from a sample using the above described oil body proteins and in some cases, ligands. In the method of the invention, oil bodies are mixed with a sample containing the desired target and the interaction between the ligand and target results in the non covalent association of the target with the oil body. Following centrifugation, the oil bodies and affinity-bound target are separated from the aqueous phase, effectively purifying the target from any contaminants present in the original sample. Repeating the washing step ensures that any remaining contaminants are removed.

Following their attachment to oil bodies, targets may be eluted under conditions determined empirically for each individual ligand-target pair. Treatment of the bound matrix with the appropriate eluent and centrifugation enables recovery of the purified target in the aqueous phase. If the target is a ligand-protein fusion containing a protease recognition site, then it may be treated with the appropriate protease to remove the ligand. The free ligand may then be separated from the target protein by re-application of the oil body affinity matrix or through conventional protein purification methods.

The chemical and physical properties of the affinity matrix may be varied in at least two ways. Firstly, different plant species contain oil bodies with different oil compositions. For example, coconut is rich in lauric oils (C12), while erucic acid oils (C22) are abundantly present in some *Brassica* spp. Furthermore, proteins associated with the oil bodies will vary between species. Secondly, the relative amounts of oils may be modified within a particular plant species by applying breeding and genetic engineering techniques or a combination of these known to the skilled artisan. These techniques aim at altering the relative activities of enzymes controlling the metabolic pathways involved in oil synthesis. Through the application of these techniques, seeds with a sophisticated set of different oils are obtainable. For example, breeding efforts have resulted in the development of a rapeseed with a low erucic acid content (Canola) (Bestor, T. H., 1994, Dev. Genet. 15: 458) and plant lines with oils with alterations in

- 24 -

the position and number of double bonds, variation in fatty acid chain length and the introduction of desirable functional groups have all been generated through genetic engineering (Töpfer et al., 1995, Science, 268: 681-685). Using similar approaches a person skilled in the art will be able to further expand on the presently available sources of oil bodies. Variant oil compositions will result in variant physical and chemical properties of the oil body fraction. Thus by selecting oilseeds or mixtures thereof from different species or plant lines as a source for oil bodies, a broad repertoire of oil body matrices with different textures and viscosities may be acquired.

10 **Applications of Oil Body Affinity Matrices**

Given that it is possible to engineer oil body affinity matrices for several classes of proteins, multiple uses for oil body based affinity matrices are envisioned. Bacteria, fungi, plants and animals all contain proteins which are able to specifically interact with agents such as ions, metals, nucleic acids, sugars, lipids and other proteins. These agents may be immobilized using oil body technology.

The oil body protein affinity matrices can be used to isolate any target molecule that can bind to the oil body protein, either directly or indirectly through a ligand molecule. Examples of target molecules that may be isolated from a sample using the methodology of the present invention include proteins, peptides, organic molecules, lipids, carbohydrates, nucleic acids, cells, cell fragments, viruses and metals. In particular, the inventors have shown that the affinity matrix of the present invention can be used to separate therapeutic proteins (such as thrombin), antibodies, metals (such as cadmium), carbohydrates (such as cellulose), organic molecules (such as biotin) and cells (such as bacterial cells).

Oil body affinity matrices may also be used to separate cells of industrial or medical interest from a mixed population of cells. For example haematopoietic stem cells, which are a subpopulation of blood cells and are used in bone marrow transplantations and in stem cell gene therapies, may be separated from other blood cells using oil body based affinity technology. In recombinant DNA technology it is often required that cells in which recombinant DNA has been successfully introduced, known as transformed cells, are distinguished and separated from cells which failed to

acquire recombinant DNA. Provided that part of the recombinant DNA expresses a cell surface protein which is complementary to a oil body based affinity ligand, it is possible to utilize oil bodies to separate transformed cells from untransformed cells. Oil body affinity technology may also be used to
5 separate cellular organelles such as chloroplasts and mitochondria from other cellular material. Viral particles may also be separated from complex mixtures.

It is also possible to immobilize a class of proteins known as metalloproteins, which contain prosthetic groups that specifically bind ions.
10 Examples of metalloproteins are haemoglobin, which binds iron, parvalbumin which binds calcium and metallothionein a protein which binds zinc and other metal ions. It is envisioned that oil bodies could be used to scavenge metals from streams of flowing material, which might be water contaminated with the waste of metals from laboratories and
15 industrial processes. Example 4 given below further illustrates this application. Other examples where proteins may be bioimmobilized and employed in a bioremediation strategy include the removal of phosphates, nitrates and phenols from waste streams. In part this approach may overcome the real or perceived limitations of bacterial bioremediation. In
20 certain instances it may not be practical or necessary to rely on affinity partitioning technology to separate the oil body matrix from the target compound. In these instances, it is envisioned that oil bodies may be immobilized on a solid inert surface which could be a flat surface or the surface of a column. A solution containing the affinity ligand may then be
25 passed over the surface coated with immobilized oil bodies whereupon selective affinity binding occurs. It is envisioned that immobilized oil bodies may be used in pipes and in ponds to assist in bioremediation.

Oil body affinity matrices can be used to isolate a recombinant polypeptide from cells. In such a case the recombinant polypeptide (i.e. the
30 target molecule) can associate either directly or indirectly with the oil bodies. This embodiment of the present invention is particularly advantageous as it allows the rapid and inexpensive manufacture of valuable, recombinantly expressed polypeptides.

Accordingly, the present invention provides a method for the isolation of a recombinant polypeptide from a cell, said cell comprising oil bodies and the recombinant polypeptide, said method comprising:

- 5 (1) contacting (i) said oil bodies with (ii) said recombinant polypeptide to allow said recombinant polypeptide to associate with said oil bodies; and
- (2) isolating said oil bodies associated with said recombinant polypeptide.

10 In accordance with the present invention, the cell may be any cell comprising oil bodies and a recombinantly expressed polypeptide. Preferably a plurality of cells is used. Suitable cells in accordance with the present invention include any animal cell, plant cell, fungal cell, yeast cell (Teher, R. et al., 1994, Yeast 10: 1421-1428), bacterial cell (Pieper-Furst et al., 1994, J. Bacteriol. 176: 4328-4337) or algae cell (Rossler, P.G., 1988, J. Physiol. 15 (London), 394-400) comprising oil bodies and a recombinantly expressed polypeptide. Preferably however plant cells are used and more preferably plant seed cells.

The recombinant polypeptide may be any polypeptide which is recombinantly expressed by the cell. The polypeptides that may be used 20 in accordance with the present invention may comprise a signal sequence that allows the direction of the polypeptide to a selected sub cellular compartment. Signal sequences that may be used include for example endoplasmatic reticulum retention signals, apoplast targeting sequences, for example the signal sequence from the tobacco pathogenesis relating sequence (PR-S) as described by Sijmons et al., (1990, Bio/Technology 8: 217-221) and other art recognized signal sequences (e.g. see: Biochemistry & Molecular Biology of Plants (2000) Buchanan, Gruissem Jones ed ISBN 0-943088-37-2). By the term "recombinantly expressed" it is meant that a 25 nucleic acid sequence encoding the polypeptide is introduced into the cell in such a manner that the cell is capable of producing the polypeptide encoded 30 by the nucleic acid sequence. Methodologies for recombinantly expressing polypeptides have hereinfure been described and are generally art-recognized (see for example: Sambrook et al., 1990, Molecular Cloning, 2nd ed., Cold Spring Harbor Press; Owen, M.R.L. and Pen, J., 1996, Transgenic

Plants: A Production System for Industrial and Pharmaceutical Proteins, John Wiley & Sons Ltd.)

5 In a preferred embodiment, the recombinant polypeptide target molecule associates with the oil bodies indirectly through a ligand molecule. The ligand molecule may be any molecule capable of associating with an oil body and the recombinant polypeptide. In preferred
10 embodiments one ligand is used, however two or more ligands may be used if desired. Where more than one ligand is used, ligands are selected to associate with each other. The ligand may associate with the oil body and with the recombinant polypeptide through non-covalent interactions, for
15 example by using a bivalent ligand. The ligand may be also be covalently linked to the recombinant polypeptide or to the oil body. In particularly preferred embodiments of the invention, the ligand molecule is a polypeptide. Accordingly, the present invention further provides a method
of isolating a recombinant polypeptide from a cell comprising oil bodies, said method comprising:

20 a) introducing into said cell (i) a first nucleic acid sequence molecule encoding a recombinant polypeptide and (ii) a second nucleic acid sequence encoding a ligand capable of associating with said recombinant polypeptide and with said oil bodies;

b) growing said cell under conditions permitting the expression of said recombinant polypeptide and said ligand;

25 c) contacting (i) said oil bodies with (ii) said recombinant polypeptide to allow said recombinant polypeptide to associate with said oil bodies through said ligand; and

d) isolating said oil bodies associated with said recombinant polypeptide.

30 In embodiments where the ligand is a polypeptide, the ligand polypeptide may conveniently be prepared as a fusion protein with the target recombinant polypeptide and recombinantly expressed in the cell. In such an embodiment, the ligand is any molecule that can bind to, or associate with, the oil bodies or oil body protein but is preferably not a protein that is normally associated with the oil bodies. The term "a protein that is normally associated with the oil bodies" includes proteins that are

normally associated with oil bodies in non-transformed or normal cells such as oil body proteins (for example oleosins) or proteins that are naturally present in normal or non-transformed cells and may associate with oil bodies when oil bodies are purified.

5 In order to prepare the fusion protein, a chimeric nucleic acid sequence is prepared that comprises (a) a nucleic acid sequence encoding the recombinant polypeptide linked to (b) a nucleic acid sequence encoding the ligand. The fusion protein may comprise a cleavage site, for example a chemical or enzymatic cleavage site, that allows for the separation of the
10 target from the ligand molecule. Preferred protein ligand molecules in accordance with the present embodiment include antibodies and fragments thereof (i.e. Fab, F(ab')₂, monoclonal antibodies, single chain antibodies, recombinantly produced binding partners). However, any protein or peptide capable of associating with the oil body and the recombinant polypeptide may be used.
15

 By "contacting the oil bodies with the recombinant polypeptide", it is meant that the oil bodies are brought into proximity of the recombinant polypeptide in a manner that allows the recombinant polypeptide to associate with the oil bodies. In one embodiment, contacting
20 of the recombinant polypeptide and the oil body is accomplished following the application of a technique resulting in the substantial disruption the cell's integrity. Generally any technique that substantially releases the cell's constituents may be used, however the technique typically varies depending on the cell type that is selected. Techniques to disrupt cells include physical
25 techniques such as the application of high pressure, as well chemical and biochemical techniques such as the use of enzymes capable of degrading cellular membranes and other art-recognized techniques. Preferably the techniques and conditions to substantially disrupt the cell's integrity are selected such that the cell's constituents are released while the oil bodies
30 remain substantially intact. In embodiments of the invention where plant seeds are used, grinding equipment such as mills, for example colloid mills, disk mills, pin mills, IKA mills, flaking rolls and orbital mills may conveniently be used. Preferably the plant seeds are ground in the presence of an aqueous solution, for example water. The embodiment of the present

invention in which the recombinant polypeptide associates with the oil body after substantially disrupting the cell is particularly desirable because it allows expression of the recombinant polypeptide in a sub-cellular compartment, for example the golgi complex, endoplasmatic reticulum or
5 apoplast, having the most appropriate physico-chemical conditions for the selected recombinant polypeptide while taking advantage of the oil body matrix as an extremely cost-effective purification tool.

In another embodiment of the invention, contacting of the recombinant polypeptide and oil bodies is accomplished within the cell. In
10 such an embodiment the recombinant polypeptide is expressed in a manner that allows the recombinant polypeptide to be directed intracellularly to the oil bodies. This could involve the expression of the recombinant polypeptide in a manner that allows the polypeptide to accumulate in the cytoplasm.

By the term "isolating" it is meant that the oil bodies associated
15 with the recombinant polypeptide are separated from other cellular constituents. The degree of purity may vary and generally depends on the desired purity of the recombinant polypeptide. In general, separation of the oil bodies may be performed as hereinbefore described. The recombinant polypeptide may further be separated from the oil body as hereinbefore
20 described. In embodiments of the invention where the recombinant polypeptide is fused to a protein-ligand and the fusion protein comprises a cleavage site, the recombinant polypeptide and the ligand may be separated by performing a cleavage reaction. Such a reaction may be performed while the fusion protein is associated with the oil body or upon prior separation of
25 the fusion protein from the oil body. An essentially pure recombinant polypeptide may be obtained using additional purification tools such as for example column chromatography.

The following examples illustrate various systems in which oil
bodies can be used as affinity matrices. It is understood that the examples
30 given below are intended to be illustrative rather than limiting.

EXAMPLES

EXAMPLE 1

Purification of Thrombin

The following example demonstrates the utility of an oil body
5 affinity matrix for the purification of thrombin. Thrombin is a serine
protease which plays a central role in blood coagulation. It cleaves
fibrinogen to produce fibrin monomers which polymerize to form the basis
of a blood clot (Fenton 1981; *Ann. N.Y. Acad. Sci.* 370: 468-495).
Alfa-thrombin consists of two polypeptide chains of 36 (A-chain) and 259
10 (B-chain) residues linked by a disulphide bridge. Degen *et al.* 1983;
Biochemistry 22: 2087-2097). Hirudin, which is found in the salivary glands of
the medicinal leech *Hirudo medicinalis*, is a very specific and potent inhibitor
of thrombin. This inhibition is a result of the non-covalent binding of
hirudin to specific parts of the alfa-thrombin chain. (Stone and Hofsteenge
15 1986; *Biochemistry* 25: 4622-4628).

The immobilized ligand is comprised of an isoform of hirudin
fused to the 18 kDa *Arabidopsis* oleosin (oil body protein) (Van Rooijen *et al.*,
1992; *Plant Mol. Biol.* 18: 1177-1179). Expression of the construct is regulated
by the *Arabidopsis* 18 kDa oleosin promoter (Van Rooijen *et al.*, 1994; *Plant*
20 *Mol. Biol.* 18: 1177-1179). The sequence of the oleosin-hirudin fusion is
shown in Figure 2 and in SEQ.ID.NO:3.

Oleosin-Hirudin Construct

Oligonucleotide primers were designed based upon the
reported sequence for a *Brassica napus* oleosin gene (Murphy *et al.* 1991,
25 *Biochim. Biophys. Acta* 1088: 86-94) and used to amplify a fragment from *B.*
napus genomic DNA through PCR. Using this fragment as a probe, a clone
carrying a 15 kbp insert was identified and isolated from a FMBL3
Arabidopsis genomic library. Oligonucleotide primers were used to amplify
a fragment from this insert containing the entire oleosin coding sequence
30 and intron together with 840 basepairs of the 5' upstream region. The
primers were designed so as to eliminate the translational stop codon and to
introduce a *Pst*I restriction endonuclease recognition site at the 5' end and a
Sall followed by a *Pvu*I site at the 3' end of the fragment. The fragment was
end-filled and ligated into the *Sma*I site of the plasmid vector pUC19. A *Sall*

- *EcoRI* fragment from plasmid pBI121 (Clontech) comprising the nopaline synthetase terminator sequence was then inserted to generate pOBILT.

5 A synthetic hirudin variant 2 (HV2) sequence was synthesized based upon reported sequence information (Harvey *et al.* 1986, *Proc. Natl. Acad. Sci. USA* 83: 1084-1088) but employing *B. napus* and *Arabidopsis* codon usage. The sequence was amplified using four overlapping oligonucleotide primers designed such that the resulting fragment possessed *PvuI* and *Sall* sites at the 5' and 3' ends respectively. This fragment was ligated into the *SmaI* site of the pUC19 plasmid vector to generate pIIR. The *PvuI* - *Sall* 10 fragment from pHIR was then inserted into pUCOBILT between the oleosin and terminator sequences to form an in-frame fusion with the oleosin coding region giving pUCOBHIRT. The entire construct was subcloned into pBluescript KS+ (pBIOBHIRT) and then into the *PstI* site of pCGN1559 plasmid (McBride and Summerfelt, 1990, *Plant Mol. Biol.* 14: 269-276) carrying 15 a neomycin phosphotransferase gene under control of the 35-S CaMV promoter (pCGBHIRT). This plasmid was introduced into *Agrobacterium tumefaciens*. The preparation of this plasmid is shown in Figure 3.

Transformation and Regeneration

20 Procedures for the transformation of *Agrobacterium* and plants have been described previously. *Agrobacterium tumefaciens* was transformed with the above construct through electroporation (Dower *et al.*, 1988; *Nucl. Acids Res.* 16: 6127-6145). The transformed bacteria were then used to transform cotyledonary explants of *Brassica napus*, followed by plant regeneration according to the methods of Moloney *et al.* (1989; 25 *Plant Cell Reports* 8: 238-242). Transgenic plant were initially identified using a neomycin phosphotransferase assay and subsequently confirmed by expression of the oleosin-hirudin fusion as determined through northern and immunoblot analysis.

Preparation of Oil Bodies

30 Seed from either control (non-transgenic) plants or transgenic plants expressing the oleosin-hirudin fusion were homogenized in five volumes of cold grinding buffer (50 mM Tris-HCl, pH 7.5, 0.4 M sucrose and 0.5 M NaCl) using a polytron operating at high-speed. The homogenate was centrifuged at approximately 10 x g for 30 min. to remove particulate

matter and to separate oil bodies from the aqueous phase containing the bulk of soluble seed protein. Oil bodies were skimmed from the surface of the supernatant with a metal spatula and placed in one volume of fresh grinding buffer. To achieve efficient washing in subsequent steps, it was important to ensure that the oil bodies were thoroughly redispersed. This was accomplished by gently re-homogenising the oil bodies in grinding buffer with the polytron operating at low-speed. Using a syringe, the resuspended oil bodies were carefully layered underneath five volumes of cold 50 mM Tris-HCl, pH 7.5 and centrifuged as above. Following centrifugation, the oil bodies were again removed and the washing procedure repeated three times to remove residual contaminating soluble seed proteins. The final washed oil body preparation was resuspended in one volume of cold 50 mM Tris-HCl pH 7.5, redispersed with the polytron, and was then ready for use as an affinity matrix.

15 Affinity Purification of Thrombin

The purification of thrombin using the oleosin-hirudin fusion protein is shown schematically in Figure 4. In order to evaluate the binding of thrombin, affinity matrices were prepared from transgenic *Brassica napus* seeds expressing the oleosin hirudin fusion protein (4A4 seeds) (Parmenter *et al.* *Plant Molecular Biology* (1995) 29: 1167-1180) and from wild type *Brassica napus* cv Westar seeds. Binding of thrombin to both matrices was evaluated. Procedures for the preparation of washed oil bodies from seeds were the same as those described above. Solutions containing a range of thrombin activities between 0 and 1 units were mixed with 10 µl of a fixed amount of affinity matrix (prepared from a total of 10 mg of dried seeds; corresponding to approximately 100 µg of total oil body protein) in 500 µl binding buffer (50 mM Tris-HCl (pH 7.5); 0.1% (w/v) BSA). The oil body suspension was then incubated for 30 minutes on ice and centrifuged at 14,000 rpm for 15 minutes at 4°C. The buffer under the oil bodies (termed 'unternatant') containing the unbound, free thrombin was recovered using an hypodermic needle and assayed for thrombin activity as follows. A total of 250 µl of unternatant was added to 700 µl binding buffer and prewarmed to 37°C. Following the addition of 50 µl of 1 mM thrombin substrate N-p-tosyl-gly-pro-arg-p-nitroanilide (Sigma) to the unternatant, the change

in optical density at 405 nanometers was monitored spectrophotometrically for 3 minutes. The concentration of thrombin in the assay mixture was determined employing a standard curve which was constructed using a set of thrombin samples containing known concentrations of thrombin. The values obtained from these assays were used to calculate the concentration bound thrombin assuming:

$$[\text{bound thrombin}] = [\text{total thrombin}] - [\text{free thrombin}]$$

The ratio of the concentration of bound over the concentration of free thrombin was plotted as a function of the concentration of bound thrombin (Scatchard plot). From these plots the dissociation constants of the affinity matrix were calculated following standard procedures (Scatchard, G. *Ann. N.Y. Acad. Sci.* (1949) 57: 660-672) and assuming: $K_a = 1/K_d$. The dissociation constants of the affinity matrices were $3.22 \times 10^{-7} \text{m}$ for wild type and $2.60 \times 10^{-6} \text{m}$ for 4A4 oil bodies.

In order to evaluate the recovery of bound thrombin from the matrices a NaCl gradient was employed. The elution profile of thrombin bound to oleosin-hirudin oil body matrices was compared with the profile from thrombin bound to wildtype oil body matrices. Procedures for preparation of wild type oil bodies from wild type *Brassica napus* cv Westar seeds and for the preparation of oleosin-hirudin oil bodies from *Brassica napus* 4A4 seeds (Parmenter et al. *Plant Molecular Biology* (1995) 29: 1167-1180) were identical to those described above. Procedures for binding of thrombin to the matrices were as described above, except 100 μl aliquots of oil bodies were used to bind 0.5 units of thrombin. Oil body suspensions were left on ice for 30 minutes prior to centrifugation for 15 minutes at 4°C and 14,000 rpm. The supernatant was assayed for (unbound) thrombin activity. The oil body matrix was then resuspended in binding buffer to which NaCl was added to a final concentration of 0.05 M. Starting with the 30 minutes incubation of the oil body suspension on ice, the procedure was repeated five times increasing the NaCl concentration in a stepwise fashion. The final NaCl concentrations used were 0.05 M, 0.1 M, 0.2 M, 0.3 M, 0.4 M and 0.6 M. The NaCl concentrations in the thrombin assay were kept constant at 150 mM. Figure 5 shows the elution profiles obtained when wildtype oil bodies and 4A4 oil bodies were used.

EXAMPLE 2

Use of Antibodies as Bivalent Ligands

Antibodies may be used as bivalent ligands by virtue of their affinity both for specific epitopes and for other antibodies or proteins (for example the *Staphylococcus aureus* protein A) which have affinity for immunoglobulins (IgGs). In this example, polyclonal anti-oleosin antibodies serve as a bivalent ligand and antibodies raised in rabbits against the anti-oleosin antibodies serve as the target. This example is illustrated schematically in Figure 6.

Oil bodies were prepared from 5 g of wild type *Brassica napus* cv Westar seeds following the procedure described in Example 1. Subsequently, oil bodies were washed twice with 100 mM glycine (pH 2.5), neutralized through two washes in binding buffer (50 mM Tris-HCl, pH 7.5) and resuspended in 5 ml of binding buffer. A 150 μ l aliquot of the washed oil body preparation was combined with 500 μ l of rabbit serum containing anti-oleosin antibodies (ligand antibodies), diluted 1:10 with binding buffer. The oil body suspension was mixed thoroughly and incubated for 1 h at 4°C with agitation. Following incubation, unbound ligand antibodies were removed from the oil body suspension through three washes with 1 ml of binding buffer. Oil bodies were then combined with 500 μ l of serum diluted 1:500 in binding buffer and containing anti-rabbit IgG antibodies (the target antibodies) conjugated with horseradish peroxidase (HRP) as a detection label (Sigma). This suspension was mixed and incubated under conditions identical to those used for the anti-oleosin antibody binding. As a control, target antibodies were incubated with oil bodies which had not been previously bound to ligand antibodies. Both samples were subsequently washed four times with 1 ml of binding buffer to remove unbound antibodies. Using binding buffer, the samples were equalized with respect to concentration of oil bodies as determined by measuring sample turbidity spectrophotometrically at 600 nm. To assay for bound target antibody, samples containing 5 μ l of oil bodies were mixed with 1 ml of the HRP colorimetric substrate tetramethylbenzidine in 0.01% hydrogen peroxide and reacted for 10 minutes at room temperature. Reactions were stopped by the addition of 500 μ l of 1 M H₂SO₄ and the absorbance at 450 nm was

determined. Corrections for the presence of residual, unbound target antibody remaining after washing were made by assaying 5 µl of the final wash fraction. The results obtained for control and ligand bound oil body preparations are set forth in Figure 7.

5 **EXAMPLE 3**

Use of Oleosin-Specific Ligands

The use of an oleosin-specific ligand represents an alternative to the use of an antibody or genetically-engineered oleosin fusion proteins for the purification of recombinant target proteins. In this case, the target
10 protein is fused to the oleosin-specific ligand and the endogenous oleosins present on the oil bodies of non-transgenic seeds serve as the complementary ligand-affinity matrix. In addition to eliminating the requirement for a transgenic line expressing an oleosin fusion, this approach increases the overall capacity of the affinity matrix, since all of the
15 endogenous oleosins may now participate in binding.

Oleosin-specific ligands may be identified and isolated from a peptide phage display library screened with oleosin protein. Since the extreme hydrophobicity of the oleosin central domain can result in aggregation and precipitation of the protein when removed from oil bodies,
20 a mutant protein lacking this domain may be used for screening. This has little effect on the efficacy of the ligand, as only the hydrophilic portions of the oleosin are exposed to the cytoplasm (i.e. the N- and C-termini). Hence, these are the only regions available for binding to a ligand. Once isolated, the ligand may be fused to a common reporter protein, green fluorescent
25 protein (GFP) (Fraser, 1995, *Trends Genet.* 11:320-323), to demonstrate purification.

Removal of the Oleosin Central Domain

Oligonucleotide primers specific for the *Arabidopsis* oleosin gene described above can be used to amplify an oleosin gene from a *B.*
30 *napus* cDNA library (van Rooijen 1993, *Ph.D. Thesis*, University of Calgary). Primers flanking sequences encoding the N-terminal 62 amino acids and the C-terminal 55 amino acids, may be used to amplify sequences for the respective N- and C-terminal oleosin domains in separate reactions. Additionally, the primer for the 5' end of the N-terminal domain contains a

sequence for a thrombin recognition site to enable cleavage of the fusion protein as described below. The resulting fragment was ligated into the *Sma*I site of the bacterial expression vector pEZZ 18 (Pharmacia). This vector contains sequences encoding a signal peptide for protein secretion into the periplasm, and synthetic IgG binding domains derived from protein A to facilitate protein purification, downstream of the multiple cloning site.

Expression and Purification of the Oleosin Deletion Construct

The vector carrying the deletion mutant construct is introduced into *E. coli* using standard methods and transformants selected. A culture of the transformed bacteria can be induced to express the synthetic protein A-mutant oleosin fusion protein by addition of 1 mM IPTG. Induced cells may be pelleted and resuspended in 5 mM MgSO₄ causing lysis of the periplasmic membrane through osmotic shock. The lysed cells are centrifuged and the supernatant containing the secreted protein is loaded on to a column containing IgG-coupled sepharose. After washing to remove unbound protein, the column is loaded with a buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl and 1.0 IU/ml of purified Bovine thrombin (Sigma) to cleave the mutant oleosin from the synthetic protein A. Following incubation at 37°C for 4h, the column is drained and the eluate passed through a column of heparin-coupled sepharose to remove thrombin. The eluate from this column, containing the mutant oleosin protein, is recovered and purity of the protein examined through gel electrophoresis followed by staining with Coomassie blue R250.

Generation of a Peptide Combinatorial Library

A random peptide combinatorial library may be generated according to the methods of Scott and Smith (1990; *Science* 249: 386-390). Briefly, the PCR is used to amplify a synthetic DNA fragment containing the degenerate sequence (NNK)₆; where 'N' represents an equal mixture of deoxynucleotides G, A, T, and C, and K represents an equal mixture of deoxynucleotides G and T. The degenerate sequence encodes for hexameric peptides among which are represented every possible combination of the 20 amino acids and amber stop codon. The PCR product is ligated into the gene III sequence of the filamentous bacteriophage fUSE and the resulting phagemid introduced into *E. coli* through electroporation.

Identification and Isolation of Oleosin-Specific Ligands

The peptide phage display libraries are amplified, concentrated and stored in aliquots of 10^{12} tdu/ml. Purified mutant oleosin protein is biotinylated using a thiol-cleavable linker (S-S biotin, Pierce) and purified by size exclusion chromatography. Aliquots of the peptide phage display library containing 5×10^{11} tdu in two ml are screened with the biotinylated protein at a concentration of 50 nM. Phage binding the mutant oleosin protein are recovered using streptavidin-coated paramagnetic beads. Following washing, the phage are eluted through the addition of 50mM dithiothreitol which cleaves the disulphide bond. The eluted phage are then incubated with an excess of log-phase F+ *E. coli*. Aliquots of the infected cells are plated to determine the phage titre and the remaining cells used in successive rounds of amplification and screening. Following enrichment of the eluted phage by 3-4 orders of magnitude, individual phage are selected and tested for binding to mutant oleosin by direct ELISA. Binding by phage is detected using anti-phage antibodies (Crosby and Schorr, 1995, In *Annual Review of Cell Biology*). Single stranded DNA is isolated from phage exhibiting binding and the peptide-encoding sequence determined.

Affinity Purification with Oleosin-Specific Ligands

The sequence for an oleosin ligand isolated as described above is fused in-frame upstream the sequence for *gfp10* (Prasher *et al.*, 1992, *Gene* 111: 229-233) encoding GFP and the construct ligated into the bacterial expression vector pKK233 (Pharmacia). Soluble protein is extracted through sonication of cells induced to express the ligand-GFP fusion, and adjusted to a concentration of 10 mg/ml in 50 mM Tris-HCl, pH 7.5.

Twenty ml of the protein solution is mixed with 2ml of oil bodies prepared as described above, from seeds of non-transgenic plants. The mixture is incubated at 4°C for 30 min with agitation to allow binding and then centrifuged to separate the oil bodies and soluble fraction. The amount of GFP remaining in the soluble fraction after removal of oil bodies is determined by fluorescence spectrofluorometry at a wavelength of 508 nm and compared with that in the original bacterial extract. The amount of bound GFP is calculated to determine the capacity of the matrix.

amount of cells present in the cell pellet as the concentration of anti-IgGs present in the oil-body *S. aureus* mixture increases.

Differential Binding of Two Strains of *Staphylococcus aureus*.

In this experiment an oil body affinity matrix is employed to demonstrate differential binding of two strains of *Staphylococcus aureus*. Formalin fixed *S. aureus* strains, one expressing the IgG binding surface antigen protein A and one lacking protein A, are commercially available from Sigma. Dilute aliquots of both *S. aureus* strains of equal OD₅₅₀ could be prepared. To each of these aliquots control oil bodies from untransformed plants or oil bodies mixed with anti-oleosin antibodies could be added. Following incubation for an appropriate length of time at an appropriate temperature, the samples could be centrifuged to pellet unbound bacterial cells and to separate the oil body fraction. The oil bodies could be decanted, vortexed and the OD₅₅₀ could be determined. The pellets could be resuspended and the OD₅₅₀ of the supernatant could be determined. It is anticipated that only in the sample containing the *S. aureus* strain expressing protein A and the oil body complexed with anti-oleosin antibodies, fractionation of these cells to the oil body fraction will be observed. Binding of the cells to the oil body could be further demonstrated by lowering of the pH of the oil body fraction. Subsequent to centrifugation the release of cells from the oil bodies could be evidenced by the presence of a pellet and/or an increase in OD₅₅₀ upon resuspension of the pellet.

Separation of *Staphylococcus aureus* from *E. coli*

A viable *S. aureus* strain could be mixed with varying quantities of cells of an *E. coli* strain having a specific antibiotic resistance. The mixed bacterial sample could be vortexed with control antibodies and with oil bodies which have been complexed with anti-oleosin antibodies. After incubation for an appropriate length of time and at an appropriate temperature oil bodies could be washed and the supernatant and oil bodies could be directly titrated and selectively plated on blood agar for *S. aureus* growth and on LB plates for *E. coli* growth. The enrichment or actual separation obtained could be determined by an estimate of colony forming units.

EXAMPLE 5

Separation of Whole Cells

The following example illustrates the capacity of oil bodies to immobilize whole cells. One potential for the use of bacterial cell separation lies in the utility for diagnostics. It is also desirable to separate unique eukaryotic cells such as lymphocytes and stem cells from complex mixtures of cells where the cell type of interest is present in relatively low numbers.

Binding of *Staphylococcus aureus* to oil bodies via protein A

For the purpose of this example, *S. aureus* cells, which express protein A as a surface antigen were mixed with oil bodies with varying amounts of polyclonal anti-oleosin antibodies.

Preparation of oil bodies

Seeds of *B. napus* cv Westar were surface sterilized in bleach, rinsed and ground with a mortar and pestle in grinding buffer (50 mM Tris pH 7.5, 0.4 M sucrose and 100 mM glycine). The homogenate was filtered through Miracloth into sterile 15 ml Corex tubes. The filtered homogenate was then centrifuged for at 4°C for 10 min at 10,000 xg. The oil body fraction was removed and resuspended in 50 mM Tris pH 7.5 and 0.4 M sucrose and washed two times using the same buffer. Aliquots of 1 ml oil bodies were transferred to 1.5 ml Eppendorf tubes and centrifuged at room temperature for 10 min at 16,000 xg. The oil bodies were washed in 50 mM Tris pH 7.5 and 0.4 M sucrose 5-6 more times until no visible pellet was observed.

Binding of *S. aureus* cells to anti-oleosin coated oil bodies

Formalin fixed *S. aureus* cells (Sigma, P-7155) were washed 3-4 times in 50 mM Tris-Cl pH 7.5. and resuspended. Washed oil bodies (300 µl) and *S. aureus* cells were mixed with varying amounts of anti-oleosin IgGs (50 µl). After mixing and incubating at room temperature for 2 hrs, the mixtures were centrifuged at room temp at 16,000 xg for 5 min. The oil body fraction and supernatant were carefully removed and the cell pellet was washed twice in 1 ml 50 mM Tris-Cl pH 7.5. The walls of the tube were wiped with a tissue to remove traces of oil. Subsequently the drained cell pellets were resuspended in 1 ml of water and the OD₆₀₀ were determined.

Figure 12 is a representative experiment showing the decrease in the

was digested with *KpnI* to release the oleP-oleMT-ubi3' insert. This expression cassette was inserted at the *KpnI* site of the binary vector pCGN1559 to yield the final construct pBIOOM3'. The sequence of the oleosin-metallothionein fusion is shown in Figure 8 and SEQ.ID.NO.6. The construction of plasmid pB100M3' is shown in Figure 9.

Transformation and Regeneration

Transgenic *B. carinata* plants expressing the oleosin-metallothionein fusion were created using transformation and regeneration protocols as described in Example 1.

Oil Body Preparation

Washed oil bodies were prepared from *B. carinata* seeds of transgenic and control plants as described in Example 1.

Removal of Cd++ From Solution Using an Oil Body Affinity Matrix

The use of the oleosin-metallothionein fusion to bind cadmium ions in solution is shown schematically in Figure 10.

A solution of 10 μ M CdCl₂ in 10 mM Tris-HCl, pH 7.2 containing 0.01 μ Ci/ml ¹⁰⁹Cd was prepared. A 1 ml aliquot of this CdCl₂ solution was thoroughly mixed with 100 μ l of washed oil bodies (1.6 mg oil body protein) prepared from seeds expressing the oleosin-metallothionein fusion protein and incubated at 22°C for 1 hr. Following centrifugation for 5' at 10,000 xg to separate the oil bodies from the aqueous phase and 2 washes in 1 ml of 10 mM Tris-Cl, pH 7.2, the amount of ¹⁰⁹Cd++ remaining bound to oil body fraction was determined using a gamma-counter (Cobra auto-gamma, Canberra Packard, Canada). An identical experiment was performed with oil bodies from non-transgenic seeds to detect and correct for non-specific binding of Cd ions to the matrix.

Cd++ ions were eluted from the oil body metallothionein affinity matrix by mixing of the oil body fraction with 1 ml of 100 mM glycine (pH = 3.0) buffer (Pazirandeh *et al.*, 1995; *Appl. Microbiol. Biotechnol.* 43: 1112-1117). Following centrifugation for 5 min. At 10,000 xg, the oil body fraction was removed and assayed for bound Cd++ ions as above. Figure 11 shows Cd binding and elution from the affinity matrix.

The oil bodies are washed twice in 20 ml of 50 mM Tris-HCl, pH 7.5, resuspended in 2 ml of the same buffer and divided into 20 aliquots of 100 μ l. Conditions for the elution of ligand-GFP fusion protein are determined by adding 1ml of solutions ranging in pH from 2-10 and in NaCl concentration from 0-1 M to different aliquots. After mixing and incubation at 4°C for 30 min, the oil bodies are removed and the soluble fractions collected. The amount of ligand-GFP fusion protein in the soluble fraction is determined by fluorescence spectrophotometry.

EXAMPLE 4

10 Removal of Heavy Metal Ions

The following example demonstrates the utility of oil body affinity matrices for the recovery/removal of non-protein targets from complex solutions. For the purpose of this example the metallothionein/ Cd^{++} ligand pair was used. However other metal binding proteins such as phytochelatins (Rausser, 1990; *Ann. Rev. Biochem.* 59: 61-86) and metal ions including Cu^{++} and Zn^{++} could also be used.

Oleosin-Metallothionein Fusion

An oleosin gene from a *B. napus* cDNA library (van Rooijen 1993, *Ph.D. Thesis*, University of Calgary) was amplified through PCR with oligonucleotide primers designed so as to create *NotI* and *NcoI* sites at the 5' and 3' ends of the gene respectively. The resulting fragment was digested and placed into the *NotI/NcoI* sites of pGN to yield plasmid poleGN. The human metallothionein gene, *mt-II* (Varshney and Gedamu, 1984, *Gene*, 31: 135-145) was amplified using oligonucleotide primers designed to create a unique *NotI* site at the 3'-end of the gene. The resulting PCR product was subcloned into the blunt-end *EcoRV* site of pBluescript KS+ to form pBSMTC. The *mt-II* gene was then excised from this plasmid and subcloned into the *NcoI/KpnI* sites of poleGN replacing the GUS-NOS region to generate pOLEMTC. The 773 base oleosin-MI fusion of pOLEMTC was excised with *NotI* digestion and inserted into the unique *NotI* site of polePN3' between the oleosin promoter (oleP; Van Rooijen *et al.*, 1992, *Plant Mol. Biol.* 18: 1177-1179) and the *P. crispum ubi4-2* gene terminator (*ubi3'*; Kawalleck *et al.*, 1993, *Plant Mol. Biol.* 21: 673-684.) to generate pOOM3'. After the fusion was determined to be in the correct orientation, pOOM3'

Identification of Pathogens Present in Low Concentrations in a Complex Mixture

For diagnostic purposes it is often desirable to concentrate bacterial or viral pathogens which invade human or animal tissues in low numbers. An oil body affinity matrix could be used to enrich for these pathogens, so that they could subsequently be identified and characterized.

Pathogens often specifically bind to human or animal cells through the interaction with a receptor or surface protein. Oleosin could be fused to the human or animal protein ligand and recombinant oil bodies could be employed to immobilize the pathogens. Examples of the formation of protein complexes formed between proteins of human and pathogenic origins known to the prior art include: human fibrinogen or fibrin specific domains which bind to *S. aureus* protein clumping factor A (clf-A) (McDevitt *et al.* 1995; *Mol. Microbiol.* 16: 895-907); human decay accelerating factor (DAF) to which urinary and intestinal tract pathogenic *E. Coli* bind (Nowicki *et al.* 1993; *J. Of Experim. Med.* 178: 2115-2121); a human cell ligand which is expressed in the carcinoma cell line Caco-2 and which binds uniquely to the 28 kD *Klebsiella pneumoniae* fimbria protein KPT-28 (Di Maretino *et al.*, 1996; *Infect. and Immun.* 64: 2263-2266) and human cell extracellular matrix fibronectin specific domains which complex specifically with *Streptococcus pyrogenes* adhesin (protein F) (Ozeri *et al.*, 1996; *EMBO J.* 15: 989-998).

EXAMPLE 6

Separation of Small Organic Molecules

This example describes how an oil body affinity matrix may be used for the recovery/removal of small organic molecules from solution. By way of example, the small organic molecule, biotin, is purified using avidin as a ligand.

Construction of Avidin Ligands

Avidin is a protein synthesized by avian species and exhibits an extremely high affinity for biotin, a natural co-factor for many carboxylases. Preparations of purified avidin (commercially available from Sigma) can be conjugated chemically to anti-oleosin antibodies using standard procedures known to those skilled in the art. This approach would

yield a bivalent avidin ligand suitable to demonstrate affinity based removal of biotin. Alternatively, an oleosin-avidin gene fusion may be utilized. The gene encoding avidin in chicken (*Gallus gallus*) has been identified and its sequence has been determined (Beattie *et al.*, 1987, *Nucl Acids Res.* 15: 3595-3606). Based on the sequence the gene for avidin could be synthesized chemically or through the PCR and fused to the *B. napus* oleosin (van Rooijen, 1993, *Ph.D. Thesis*, University of Calgary) as described in example 4. Streptavidin, an analogous bacterial biotin binding protein, could also be employed.

10 **Oil Body Preparation**

Washed oil bodies would be prepared from seeds of transgenic plants and/or control plants as described in example 1.

Binding of Bivalent Avidin-Oleosin Ligand

15 Binding of anti-oleosin antibodies and removal of unbound ligand will be as detailed in example 3.

Removal of Biotin from Solution

20 Solutions containing known concentrations of biotin could be combined with a fixed amount of oil bodies complexed with an anti oleosin antibodies conjugated with avidin. Following binding, the mixture would be centrifuged to separate oil body and aqueous fraction. The amount of biotin remaining in the aqueous fraction is determined by competitive ELISA using anti-biotin antibodies conjugated to horse radish peroxidase (HRP). The amount of bound biotin may be calculated assuming:

$$[\text{bound biotin}] = [\text{total biotin}] - [\text{free biotin}]$$

25 From the obtained values, the dissociation constants can be determined as described in example 2. As a control, an identical experiment could be performed with oil bodies bound to anti oleosin antibodies which have not been conjugated with avidin. If desired, biotin could be released from the oil body-avidin matrix through competitive elution using an excess of 2-(4'-hydroxybenzene) benzoic acid (HABA). Elution may also be aided by employing a genetically engineered mutant of avidin which exhibits a lower affinity for biotin. Such mutants have been described for the analogous biotin binding protein from bacteria, streptavidin (Chilkoti *et al.*, 1995; *Bio/Technol.* 13: 1198-1204).

EXAMPLE 7

Separation of Carbohydrates

The following example describes the utility of oil body matrices for the recovery of carbohydrates from complex biological mixtures. In this example the inventors demonstrate that an oil body immobilized cellulase is capable of binding cellulose.

Oleosin-Cellulose Binding Domain Fusion

Several of the cellulases produced by the bacterium *Cellulomonas fimi* contain discrete cellulose binding domains (CBDs). These CBDs independently bind to cellulose even when they are separated by proteolytic cleavage or genetic manipulation from the catalytic domain of the enzyme. Plasmid pUC18-CBDPT contains a fragment coding for the CBD of the beta-1,4-glucanase (Gilkes *et al.*, 1992, *Journal of Biol. Chem.* 267: 6743-6749) and could be used to construct an oleosin-CBD gene fusion. A DNA fragment encoding the CBD domain could be isolated from pUC18-CBDPT using appropriate restriction enzymes or using the PCR. Alternatively, the CBDs of other cellulases from *C. Fimi* or cellulases from other sources could be used. An oleosin gene from *B. Napus* isolated from a cDNA library (van Rooijen, 1993, *Ph.D. Thesis*, University of Calgary) was cloned in pCGN using the PCR and yielding plasmid pOLEGN as described in example 4. An in-frame gene fusion between the oleosin gene and the CBD gene could be generated using standard molecular techniques known to those skilled in the art. The final construct would comprise the CBD domain translationally fused immediately downstream of the oleosin.

Transformation and Regeneration

In order to introduce the fusion gene construct in plants, it would be subcloned in a binary vector, such as pCGN1559. Transgenic plants which express the oleosin-CBD fusion could be generated as described in example 1.

Oil Body Preparation

Washed oil bodies could be prepared from the seeds of transgenic and control wild type plants as described in example 1.

Removal of Cellulose from Solution Using an Oil Body Affinity Matrix

In order to evaluate binding of cellulose to the oil body affinity matrix, the binding capacities of oil bodies of wild type and transgenic plants are compared. Oil bodies could be mixed with appropriately buffered solutions containing a range of cellulose concentrations. The oil body suspension could then be incubated for an appropriate length of time and at an appropriate temperature. Upon centrifugation, the supernatant could be recovered and assayed for cellulose concentrations. The concentrations bound cellulose and free cellulose could be calculated assuming:

$$[\text{bound cellulose}] = [\text{total cellulose}] - [\text{free cellulose}]$$

The ratio of the concentration bound over the concentration free cellulose could be plotted as a function of the concentration of bound cellulose. From these plots dissociation constants could be calculated following standard procedures (Scatchard, *G. Ann. N. Y. Acad. Sci.* (1949) 57: 660-672) and as detailed in example 2.

EXAMPLE 8

Separation of Nucleic Acids

The following example describes a method in which oil bodies are employed to bind single stranded (SS) nucleic acids.

Isolation of Single Stranded Nucleic Acids

A method for capturing SS nucleic acids may be used in diagnostics, such as plant viral disease, or in research applications where non-reannealed SS nucleic acids need to be selectively removed from solutions such as in hybridization reactions for differential screening of expressed genes. Oleosins could be fused with SS DNA or RNA binding proteins or specific domains thereof and could be used to trap SS nucleic acids for identification or further amplification. Well characterized SS nucleic acid binding proteins include: Agrobacterial Ti plasmid Vir E2 protein (Zupan *et al.*, 1995, *Plant Physiol.* 107: 1041-1047); Tobacco Mosaic Virus (TMV) movement protein P30 (Citovsky *et al.*, 1990; *Cell* 60: 637-647; Waigmann *et al.*, 1994 *Proc Natl. Acad. Sci (USA)* 91: 1433-1437); Cauliflower Mosaic Virus coat protein (Thompson *et al.*, 1993; *J. Gen. Virol*

74: 1141-1148) and *E. Coli* RecA and single stranded binding (SSB) proteins (Radding, 1991 *J. Biol. Chem.* 266: 5355-5358).

EXAMPLE 9

Separation of Recombinant Proteins

5 The following example further demonstrates the utility of an oil body affinity matrix for the purification of recombinant target proteins. For the purpose of this example, the IgG/protein A ligand pair has been chosen. The construct employed consists of a protein A domain which was fused to the 18 kDa *Arabidopsis* oleosin (Van Rooijen *et al.*, 1992; *Plant Mol. Biol.* 18: 1177-1179). Oil bodies containing oleosin-protein A fusion proteins were isolated and used to demonstrate specific binding of rabbit-anti-mouse IgGs conjugated to Horse Raddish Peroxidase (HRP). The configuration of the oleosin-protein A fusion on the oil body and binding of IgG to the fusion is shown in Figure 15.

15 The Oleosin-Protein A Fusion

 A synthetic protein A sequence encoding a protein capable of binding to IgG was synthesized based on reported sequence information (pRIT2T, protein A gene fusion vector; Pharmacia) and was amplified through the PCR. Each primer used in the PCR contained restriction sites 5' to the protein A-specific sequence in order to facilitate cloning. The reverse primer (i.e. the primer in the antisense direction) also contained a translational stop codon following the coding sequence. Fig 13 shows the position of the PCR primers relative to the protein A sequence. (The protein A sequence and the primer sequences are also separately shown in SEQ.ID.NO:8, SEQ.ID.NO:10 and SEQ.ID.NO:11 respectively). The resulting fragment was ligated into a pUC19 plasmid carrying the *Arabidopsis* oleosin gene comprised of an 867 bp upstream promoter region followed by the coding region (with its associated intron) from which the translational stop codon had been removed. The 3' end of the construct contains the nopaline synthase transcriptional terminator. A spacer sequence encoding a recognition sequence for the endoprotease thrombin was incorporated immediately downstream of the oleosin coding sequence. The protein A gene sequence was introduced between this spacer sequence and the terminator sequence. In the final expression construct the oleosin and

protein A coding regions were fused in the same reading frame. The entire construct (Figure 14 and SEQ.ID.NO:12) was then excised from the pUC19 plasmid and subcloned into the plant transformation vector pCGN1559 (McBride and Summerfelt, 1990, *Plant Mol. Biol.* 14: 269-276) carrying a
5 neomycin phosphotransferase gene under the control of the 35S CaMV promoter. The resulting plasmid was introduced in *Agrobacterium* (strain EHA101).

Transformation and Regeneration

Plants were transformed and regenerated as described in
10 example 1. Transgenic plants were initially identified using a neomycin phosphotransferase assay and subsequently confirmed by expression of protein A fusions through immunoblot analysis.

Preparation of Oil Bodies

Oil bodies from the transgenic *B. napus* and *B. carinata* lines
15 expressing the oleosin-protein A fusion were prepared following the procedure described in example 1.

Binding of Oleosin-Protein A Fusions to IgG

Oil body protein extracts (20 µg/ aliquot) from various transgenic *B. napus* lines expressing oleosin-protein A fusion proteins were
20 subjected to polyacrylamide gelelectrophoresis and subsequently transferred to a PVDF membrane following standard procedures. The membrane was then probed with a HRP-conjugated mouse anti-rabbit antibody and visualised following the procedure as outlined in Antibodies, a laboratory manual (Harlow and Lane, 1988, Cold Spring Harbor). In Figure
25 16 the stained PVDF membrane is shown. A 50 kDa protein (predicted molecular mass of the oleosin-protein A fusion protein: 48,801 Da) was specifically detected in the protein extracts of all of the six transgenic *B. napus* lines tested. Untransformed control plants did not exhibit HRP activity, while the a 30 kDa protein (predicted molecular mass 29,652 Da)
30 was present in a bacterial lysate transformed with pRIT2T encoding protein A and undetectable in the untransformed lysate.

Binding and Elution of IgGs to Oil Bodies Containing Oleosin-Protein A Fusion Proteins

Washed oil bodies (10 mg/ml protein) were prepared from wildtype *B. napus* and a transgenic *B. napus* line transformed with a construct expressing an oleosin-protein A fusion protein as described in example 1 and suspended in 10 mM Tris-Cl pH 8.0. A volume of 2 μ l (\pm 34 μ g) of HRP-conjugated rabbit anti-mouse antibodies (Sigma, cat no A9044) was added to 500 μ l of the washed oil body preparation and the suspension was incubated for 1 hr at room temperature or overnight at 4°C. The samples were then centrifuged for 15 min at 16,000 xg and the supernatant was removed. Subsequently, the oil bodies were thoroughly resuspended in 500 μ l 10 mM Tris-Cl pH 8.0 using a pestle. This washing step in Tris-Cl was repeated 4 times (henceforth termed washed oil body preparation). A 5 μ l aliquot from the washed oil body preparation was washed a fifth time and then assayed for HRP activity.

HRP assays were carried out by adding 1 μ l of the washed oil body preparation to 1 ml of HRP assay mix (9.8 ml of 0.1 M NaOAc, 0.2 ml of 2.5 mg/ml 3,3',5,5'-tetramethylbenzidine in DMSO, 4 μ l H₂O₂) and incubating the mixture for 5 min at room temperature. The reaction was then stopped by adding 0.5 ml 1M H₂SO₄. The samples were filtered through a 0.22 μ m filter and subsequently the OD₄₅₀'s were determined spectrophotometrically.

In order to elute the IgGs from the oil bodies, the washed oil body preparation was resuspended in 100 mM glycine pH 3.0 and centrifuged for 15 min at 16,000 xg and incubated for 30' at room temperature. Following neutralization in 500 μ l 100 mM Tris-Cl pH 8.0, both the oil body fraction and the eluate were assayed for HRP activity as above. The binding and elution of IgGs to oil bodies from wild type *B. napus* and transgenic *B. napus* expressing an oleosin protein A fusion, are illustrated in Figure 17.

All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

EXAMPLE 10

Construction of a FRS-OBScFv-prochymosin plant expression vector

The example below describes the construction of a plant expression vector containing a gene sequence which upon expression in the plant produces a fusion protein comprising a signal sequence, a Single-chain variable Fragment (ScFv) isolated from a mouse hybridoma cell line producing monoclonal antibodies raised against an *Arabidopsis* oleosin and the zymogen chymosin.

Arabidopsis cDNA clone Atts0278 containing an *Arabidopsis* oleosin cDNA sequence was obtained from the *Arabidopsis* Biological Resource Centre (ABRC, <http://aims.cps.msu.edu>) *Arabidopsis* stock centre. This cDNA sequence is identical to the coding sequence of the *Arabidopsis* oleosin genomic sequence as published in (Van Rooijen et al (1992) Plant Mol. Biol.18: 1177-1179). Using standard molecular biology techniques such as polymerase chain reaction (PCR), ligation and transformation, this sequence was furnished with restriction sites which allowed for the in-frame cloning of this sequence in the bacterial expression vector pRSETB (obtained from Invitrogen). Manufacture's (Invitrogen's) protocols were used to express the His-tagged *Arabidopsis* oleosin. The His-tagged *Arabidopsis* oleosin protein was run on a polyacrylamide gel, electroeluted and used to obtain mouse hybridoma monoclonal antibodies using standard laboratory techniques (See eg Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989). A mouse hybridoma cell line which produces a monoclonal antibody which specifically recognizes the *Arabidopsis* oleosin described above (and very similar Brassica oleosins) was called D9.

The hybridoma cell-line was used as a source of mRNA for the production of a single-chain variable fragment (ScFv) gene in which the variable regions of the antibody heavy (Vh) and light (Vl) chain genes are joined by a flexible peptide linker. This mRNA was isolated using standard laboratory techniques. The production of the anti-oleosin ScFv gene was achieved utilizing a "Recombinant Phage Antibody System Mouse ScFv Module" obtained from Pharmacia Biotech (Code No: 27-9400-01). The OBScFv gene will be referred to as the Oil body Single-chain variable Fragment (OBScFv) gene. The OBScFv gene was furnished with an ER signal

sequence termed PRS at the 5' end of the OBScFv gene to allow for secretion into the plant cell apoplast upon expression of this gene in plants. This was accomplished using gene splicing by an overlap extension technique (Horton et al GENE (1989) 15: 61-68. The PRS DNA sequence was

5 chemically synthesized and encodes a signal sequence which is identical to the deduced amino acid sequence of a tobacco E2 thaumatin-like protein [Van Kan et al (1989) Plant Mol Biol 12: 153-155]. The 3' end of the OBScFv gene was fused in frame to a sequence encoding the bovine zymogen

10 prochymosin using gene splicing by overlap extension technique (Horton et al GENE (1989) 15: 61-68). The gene sequence encoding this prochymosin was codon optimized for expression in plants. The deduced amino acid sequence of the prochymosin is identical to the deduced amino acid sequence of prochymosin cDNA as reported in Harris et al (1982) NAR 10: 2177-2187.

The PRS-OBScFv-Prochymosin gene fusion as described above

15 was placed under the regulatory control of the phaseolin promoter and the phaseolin terminator derived from the common bean *Phaseolus vulgaris* (Slightom et al (1983) Proc. Natl Acad Sc USA 80: 1897-1901; Sengupta-Gopalan et al., (1985) PNAS USA 82: 3320-3324)). A gene splicing by overlap extension technique (Horton et al GENE (1989) 15: 61-68) was used to fuse

20 the phaseolin promoter to the PRS-OBScFv-Prochymosin gene. Standard molecular biology laboratory techniques (see eg: Sambrook et al. (1990) Molecular Cloning, 2nd ed. Cold Spring Harbor Press) were used to furnish the phaseolin promoter and terminator with Pst I and HindIII/KpnI sites respectively (see Figure1). Standard molecular biology laboratory

25 techniques were also used to place the phaseolin terminator downstream from the PRS-OBScFv-Prochymosin gene. The PstI-phaseolin promoter-PRS-OBScFv-Prochymosin-phaseolin terminator-KpnI insert sequence was cloned into the PstI KpnI sites of pSBS3000 (pSBS3000 is a derivative from the *Agrobacterium* binary plasmid pPZP221 (Hajdukiewicz et al., 1994, Plant

30 Mol. Biol. 25: 989-994). In pSBS3000, the CaMV35S promoter-gentamycin resistance gene-CAMV 35S terminator of pPZP221 was replaced with parsley ubiquitin promoter-phosphinothricin acetyl transferase gene-parsley ubiquitin termination sequence to confer resistance to the herbicide glufosinate ammonium. The resulting plasmid is called pSBS2168. The

sequence of the phaseolin promoter- PRS-OBScFv-Prochymosin-phaseolin terminator sequence is shown in Figure 1.

Plasmid pSBS168, was electroporated into *Agrobacterium* strain EHA101. This *Agrobacterium* strain was used to transform *Arabidopsis*.

- 5 *Arabidopsis* transformation was performed essentially as described in "Arabidopsis Protocols; Methods in molecular biology Vol 82. Edited by Martinez-Zapater JM and Salinas J. ISBN 0-89603-391-0 pg 259-266 (1998) except that the putative transgenic plants were selected on agarose plates containing 80µM L-phosphinothricine, and subsequently transplanted to soil
10 and allowed to set seed.

EXAMPLE 11

Polyacrylamide gelelectrophoresis (PAGE) and immunoblotting of transgenic seed extracts.

- Preparation of total Arabidopsis seed extracts for PAGE.* 50 mg of
15 transgenic *Arabidopsis* seeds were ground in 1 ml of oil body extraction buffer containing 0.4 M sucrose, 50 mM Tris-Cl pH 7.5 and 0.5 M NaCl. A 25 µl aliquot was taken out and labeled "total extract". The rest of the extract was placed at room temperature for 20-30 minutes to allow the PRS-OBScFv-Prochymosin protein to associate with the *Arabidopsis* oil bodies.
20 This mixture was spun at room temperature for 10 minutes at 10,000 x g to allow for the separation of the oil body fraction (floating on top) from the supernatant (25 µl was taken out and labeled Sup0,) and the pellet fraction. The oil body fraction was resuspended in 300 µl oil body washing buffer containing containing 0.4 M sucrose, 50 mM Tris-Cl pH 7.5. This fraction is
25 called Obtotal.

- 25 µl of Obtotal was mixed with 25 µl of a high stringency washing buffer (8M urea, 100mM Na₂CO₃) spun for 10 minutes at 16,000 x g . The oil body fraction was resuspended in 25 µl high stringency washing buffer spun for 5 minutes and the oil body fraction was resuspended in 25 µ
30 high stringency washing buffer. This fraction is called OBHighS.
The remainder of Obtotal was spun at room temperature for 10 minutes at 16,000 xg and the oil body fraction was resuspended in 300µl oil body washing buffer. A 25 µl aliquot sample was taken and this fraction is called OBO

The "total extract", Sup 0, OB0, and OBHighS samples were prepared for SDS PAGE and Western blotting using standard molecular biology protocols. Equal amounts of sample were loaded in each lane of the gel (As the total extract has a volume of 1ml and the oil bodies are resuspended in approximately 300µl, the oil body associated proteins are enriched about 3-fold in lane 4 and 5 compared to lane 2). The results of this experiment are shown in figure 2. As can be seen from this figure the majority of the OBScFvProchymosin protein fusion has been autocatalytically processed into mature chymosin and several proteolytic breakdown products of chymosin (here referred to as "partial chymosin". This autocatalytical processing is typical for secreted chymosin fusion proteins (personal observations and Ward et al (1990) Bio/Technology 8: 435-440). As seen in lane 2, in addition to mature and partial chymosin, two bands of the predicted molecular weight of an intact OBScFvProchymosin fusion protein can be found. The top band has a slightly lower mobility on a polyacrylamide gel and does not separate with *Arabidopsis* oil bodies (lane 4 and 5). Instead this band is found in the supernatant fraction (lane 3). It is predicted that this is a suboptimally folded form of the OBScFvProchymosin fusion protein. The mature and partial chymosin products are not expected to co-purify with the oil bodies as they have been cleaved from the OBScFv "carrier". As can be seen in figure 2 the ratio of OBScFvProchymosin to mature chymosin is dramatically increased in the oil body samples (lane 4 and 5) which indicates that the OBScFvProchymosin protein has a higher and specific affinity for oil bodies. When a high stringency urea buffer is used (lane 5), the only protein which still co-separates with the oil bodies is the OBScFvProchymosin fusion protein. This confirms that this binding is specific and that the OBScFv portion of the OBScFvProchymosin protein is responsible for this binding. Equal amounts of sample were loaded in each lane of the gel.

WE CLAIM:

1. A method for the separation of a target molecule from a sample comprising:
 - 1) contacting (i) oil bodies with (ii) a sample containing the target molecule to allow the target molecule to associate with the oil bodies; and
 - 2) separating the oil bodies associated with the target molecule from the sample.
2. A method according to claim 1 wherein said target molecule associates with said oil bodies through a ligand molecule that associates with the oil bodies and the target molecule.
3. A method according to claim 2 wherein the ligand molecule is covalently attached to the target molecule.
4. A method according to claim 3 wherein the target molecule is a protein.
5. A method according to claim 3 wherein the ligand molecule is a protein.
6. A method according to claim 5 wherein the protein ligand is prepared as a fusion protein with the protein target molecule and wherein the ligand is not a protein that is normally associated with oil bodies.
7. A method according to claim 6 wherein the ligand molecule is an antibody or a fragment thereof.
8. A method according to claim 7 wherein the antibody binds to an oil body protein.

9. A method according to claim 7 wherein the antibody is a single chain antibody.
10. A method according to claim 1 wherein the sample is a cell.
11. A method according to claim 2 wherein the ligand is a bivalent antibody that binds to both the oil body and the target.
12. A method according to claim 8 wherein the oil body protein is an oleosin.
13. A method according to claim 12 wherein the oleosin is derived from a plant selected from the group consisting of rapeseed (*Brassica* spp.), soybean (*Glycine max*), sunflower (*Helianthus annuus*), oil palm (*Elaeis guineensis*), coconut (*Cocos nucifera*), castor (*Ricinus communis*), safflower (*Carthamus tinctorius*), mustard (*Brassica* spp. and *Sinapis alba*), coriander (*Coriandrum sativum*) linseed/flax (*Linum usitatissimum*), thale cress (*Arabidopsis thaliana*) and maize (*Zea mays*).
14. A method according to claim 1 wherein the oil bodies associated with the target molecule are separated from the sample in step (2) by centrifugation, floatation or size exclusion.
15. A method according to claim 1, further comprising 3) separating the target molecule from the oil bodies.
16. A method according to claim 15 wherein the target molecule is separated by elution under appropriate conditions.
17. A method according to claim 1 wherein the oil bodies are obtained from the group of plants consisting of rapeseed (*Brassica* spp.), soybean (*Glycine max*), sunflower (*Helianthus annuus*), oil palm (*Elaeis guineensis*), coconut (*Cocos nucifera*), castor (*Ricinus communis*), safflower (*Carthamus tinctorius*), mustard (*Brassica* spp. and *Sinapis alba*), coriander

(*Coriandrum sativum*) linseed/flax (*Linum usitatissimum*), thale cress (*Arabidopsis thaliana*) and maize (*Zea mays*).

18. A method according to claim 1 for the isolation of a recombinant polypeptide from a cell, said cell comprising oil bodies and the recombinant polypeptide, said method comprising:

- (1) contacting (i) said oil bodies with (ii) said recombinant polypeptide to allow said recombinant polypeptide to associate with said oil bodies; and
- (2) isolating said oil bodies associated with said recombinant polypeptide.

19. A method according to claim 18 wherein said recombinant polypeptide associates with said oil bodies through a ligand that associates with the recombinant polypeptide and the oil bodies.

20. A method according to claim 19 wherein said ligand is an antibody, an antibody fragment or a single chain antibody that binds to an oil body protein.

21. A method according to claim 19 wherein the ligand is a polypeptide and said recombinant polypeptide is prepared as a fusion protein with said ligand and wherein the ligand is not a protein that is normally associated with oil bodies.

22. A method according to claim 19 comprising:

- a) introducing into said cell (i) a first nucleic acid sequence molecule encoding a recombinant polypeptide and (ii) a second nucleic acid sequence encoding a ligand capable of associating with said recombinant polypeptide and with said oil bodies;
- b) growing said cell under conditions permitting the expression of said recombinant polypeptide and said ligand;

09707167 440700

c) contacting (i) said oil bodies with (ii) said recombinant polypeptide to allow said recombinant polypeptide to associate with said oil bodies through said ligand; and

5 d) isolating said oil bodies associated with said recombinant polypeptide.

23. A method according to claim 22 wherein said recombinant polypeptide is prepared as a fusion protein with said ligand and wherein the ligand is not a protein that is normally associated with oil bodies.

10 24. A method according to claim 23 wherein said ligand is an antibody, an antibody fragment or single chain antibody that binds to an oil body protein.

25. A method according to claim 20 wherein said contacting results in the substantial disruption of the cell's integrity.

15 26. A composition comprising oil bodies associated with a ligand molecule covalently attached to a target molecule.

27. A composition according to claim 26 wherein the ligand molecule and the target molecule are proteins.

20 28. A composition according to claim 27 wherein the ligand molecule and target molecules are covalently attached as a recombinant fusion protein and wherein the ligand is not a protein that is normally associated with oil bodies.

090407-140700

ABSTRACT OF THE DISCLOSURE

A method for the separation of a target molecule from a mixture is described. The method employs oil bodies and their associated proteins as affinity matrices for the selective, non-covalent binding of
5 desired target molecules. The oil body proteins may be genetically fused to a ligand having specificity for the desired target molecule. Native oil body proteins can also be used in conjunction with an oil body protein specific ligand such as an antibody or an oil body binding protein. The method allows the separation and recovery of the desired target molecules due to
10 the difference in densities between oil bodies and aqueous solutions.

050246-140000

FIGURE 1

1 ATG GCG GAT ACA GCT AGA GGA ACC CAT CAC GAT ATC ATC GGC AGA GAC CAG TAC CCG ATG 60
 1 M A D T A R G T M M D T I G R D Q Y P M 20
 81 ATG GGC CGA GAC CGA GAC CAG TAC CAG ATG TCC GGA CGA GGA TCT GAC TAC TCC AAG TCT 120
 21 M C R D R D Q Y Q M S G R G S D Y S K S 40
 121 AGG CAG ATT GCT AAA GCT GCA ACT GCT GTC ACA GCT GGT GGT TCC CTC GTT GTT CTC TCC 180
 41 R Q I A K A A T A V T A G G S L L V L S 60
 181 AGC GTT ACC GTT GTT GGA ACT CTC ATA GCT TTG ACT GTT GCA ACA COT CTG CTC GTT ATC 240
 61 S L T L V C T V I A L T V A T P L L V I 80
 241 TTC AGC CCA ATC GTT GTC CCG GCT CTC ATC ACA GTT GCA CTC CTC ATC ACC GGT TTT GTT 300
 81 F S P I L V P A L I T V A L L I T G F L 100
 301 TCC TCT GGA GGG TTT GGC ATT GGG GCT ATA ACC GTT TTC TCT TGG ATT TAC AAG TAC GCA 360
 101 S S C G F G I A A I T V F S W I Y R Y A 120
 361 ACG GGA GAG CAC CCA CAG GGA TCA GAC AAG TTG GAC AGT GCA AGG ATG AAG TTG GGA AGC 420
 121 T G E H P Q G S D K L D S A R M R L C E 140
 421 AAA GCT CAG GAT CTG AAA CAG AGA GCT CAG TAC TAC GGA CAG CAA CAT ACT GGT GGC GAA 480
 141 K A Q D L K D R A G Y Y G Q Q H T G G E 160
 481 CAT GAC CGT GAC CGT ACT CGT GGT GGC CAG CAC ACT ACT TAA
 161 H D R D R T K G G Q H T T *

09707467-440900

FIGURE 2A

[illegible]

Variable	Mean	SD	Min	Max
Age	34.5	10.2	21	55
Gender	0.5	0.5	0	1
Marital status	0.6	0.5	0	1
Education	12.5	1.5	9	16
Income	15.2	5.8	5	35
Occupation	1.2	0.8	0	2
Health status	1.8	0.5	1	3
Stress level	2.5	0.8	1	4
Life satisfaction	3.2	0.7	2	4
Resilience	2.8	0.6	1	4
Optimism	3.5	0.9	2	4
Gratitude	3.8	0.8	2	4
Forgiveness	3.6	0.7	2	4
Empathy	3.4	0.6	2	4
Compassion	3.3	0.5	2	4
Kindness	3.7	0.7	2	4
Generosity	3.9	0.8	2	4
Patience	3.6	0.6	2	4
Self-control	3.5	0.5	2	4
Emotional stability	3.4	0.4	2	4
Conscientiousness	3.3	0.3	2	4
Openness to experience	3.2	0.2	2	4
Agreeableness	3.1	0.1	2	4
Neuroticism	2.9	0.1	2	4
Extraversion	3.0	0.1	2	4
Introversion	2.8	0.1	2	4
Shyness	2.7	0.1	2	4
Loneliness	2.6	0.1	2	4
Isolation	2.5	0.1	2	4
Depression	2.4	0.1	2	4
Anxiety	2.3	0.1	2	4
Stress	2.2	0.1	2	4
Life stress	2.1	0.1	2	4
Work stress	2.0	0.1	2	4
Relationship stress	1.9	0.1	2	4
Health stress	1.8	0.1	2	4
Financial stress	1.7	0.1	2	4
Academic stress	1.6	0.1	2	4
Social stress	1.5	0.1	2	4
Family stress	1.4	0.1	2	4
Community stress	1.3	0.1	2	4
National stress	1.2	0.1	2	4
Global stress	1.1	0.1	2	4
World stress	1.0	0.1	2	4
Universe stress	0.9	0.1	2	4
Cosmos stress	0.8	0.1	2	4
Galaxy stress	0.7	0.1	2	4
Star stress	0.6	0.1	2	4
Planet stress	0.5	0.1	2	4
Moon stress	0.4	0.1	2	4
Sun stress	0.3	0.1	2	4
Earth stress	0.2	0.1	2	4
Water stress	0.1	0.1	2	4
Fire stress	0.0	0.1	2	4
Air stress	-0.1	0.1	2	4
Earth stress	-0.2	0.1	2	4
Water stress	-0.3	0.1	2	4
Fire stress	-0.4	0.1	2	4
Air stress	-0.5	0.1	2	4
Earth stress	-0.6	0.1	2	4
Water stress	-0.7	0.1	2	4
Fire stress	-0.8	0.1	2	4
Air stress	-0.9	0.1	2	4
Earth stress	-1.0	0.1	2	4
Water stress	-1.1	0.1	2	4
Fire stress	-1.2	0.1	2	4
Air stress	-1.3	0.1	2	4
Earth stress	-1.4	0.1	2	4
Water stress	-1.5	0.1	2	4
Fire stress	-1.6	0.1	2	4
Air stress	-1.7	0.1	2	4
Earth stress	-1.8	0.1	2	4
Water stress	-1.9	0.1	2	4
Fire stress	-2.0	0.1	2	4
Air stress	-2.1	0.1	2	4
Earth stress	-2.2	0.1	2	4
Water stress	-2.3	0.1	2	4
Fire stress	-2.4	0.1	2	4
Air stress	-2.5	0.1	2	4
Earth stress	-2.6	0.1	2	4
Water stress	-2.7	0.1	2	4
Fire stress	-2.8	0.1	2	4
Air stress	-2.9	0.1	2	4
Earth stress	-3.0	0.1	2	4
Water stress	-3.1	0.1	2	4
Fire stress	-3.2			

Variable	Mean	SD	Min	Max
Age	34.5	10.2	21	55
Gender	0.5	0.5	0	1
Marital status	0.6	0.5	0	1
Education	12.5	1.5	9	16
Income	15.2	5.8	5	35
Occupation	1.2	0.8	0	2
Health status	1.8	0.5	1	3
Stress level	2.5	0.8	1	4
Life satisfaction	3.2	0.7	2	4
Resilience	2.8	0.6	1	4
Optimism	3.5	0.9	2	4
Gratitude	3.8	0.8	2	4
Forgiveness	3.6	0.7	2	4
Empathy	3.4	0.6	2	4
Compassion	3.3	0.5	2	4
Kindness	3.7	0.7	2	4
Generosity	3.9	0.8	2	4
Patience	3.6	0.6	2	4
Self-control	3.5	0.5	2	4
Emotional stability	3.4	0.4	2	4
Conscientiousness	3.3	0.3	2	4
Openness to experience	3.2	0.2	2	4
Agreeableness	3.1	0.1	2	4
Neuroticism	2.9	0.1	2	4
Extraversion	3.0	0.1	2	4
Introversion	2.8	0.1	2	4
Shyness	2.7	0.1	2	4
Loneliness	2.6	0.1	2	4
Isolation	2.5	0.1	2	4
Depression	2.4	0.1	2	4
Anxiety	2.3	0.1	2	4
Stress	2.2	0.1	2	4
Life stress	2.1	0.1	2	4
Work stress	2.0	0.1	2	4
Relationship stress	1.9	0.1	2	4
Health stress	1.8	0.1	2	4
Financial stress	1.7	0.1	2	4
Academic stress	1.6	0.1	2	4
Social stress	1.5	0.1	2	4
Family stress	1.4	0.1	2	4
Community stress	1.3	0.1	2	4
National stress	1.2	0.1	2	4
Global stress	1.1	0.1	2	4
World stress	1.0	0.1	2	4
Universe stress	0.9	0.1	2	4
Cosmos stress	0.8	0.1	2	4
Galaxy stress	0.7	0.1	2	4
Star stress	0.6	0.1	2	4
Planet stress	0.5	0.1	2	4
Moon stress	0.4	0.1	2	4
Sun stress	0.3	0.1	2	4
Earth stress	0.2	0.1	2	4
Water stress	0.1	0.1	2	4
Fire stress	0.0	0.1	2	4
Air stress	-0.1	0.1	2	4
Earth stress	-0.2	0.1	2	4
Water stress	-0.3	0.1	2	4
Fire stress	-0.4	0.1	2	4
Air stress	-0.5	0.1	2	4
Earth stress	-0.6	0.1	2	4
Water stress	-0.7	0.1	2	4
Fire stress	-0.8	0.1	2	4
Air stress	-0.9	0.1	2	4
Earth stress	-1.0	0.1	2	4
Water stress	-1.1	0.1	2	4
Fire stress	-1.2	0.1	2	4
Air stress	-1.3	0.1	2	4
Earth stress	-1.4	0.1	2	4
Water stress	-1.5	0.1	2	4
Fire stress	-1.6	0.1	2	4
Air stress	-1.7	0.1	2	4
Earth stress	-1.8	0.1	2	4
Water stress	-1.9	0.1	2	4
Fire stress	-2.0	0.1	2	4
Air stress	-2.1	0.1	2	4
Earth stress	-2.2	0.1	2	4
Water stress	-2.3	0.1	2	4
Fire stress	-2.4	0.1	2	4
Air stress	-2.5	0.1	2	4
Earth stress	-2.6	0.1	2	4
Water stress	-2.7	0.1	2	4
Fire stress	-2.8	0.1	2	4
Air stress	-2.9	0.1	2	4
Earth stress	-3.0	0.1	2	4
Water stress	-3.1	0.1	2	4
Fire stress	-3.2			

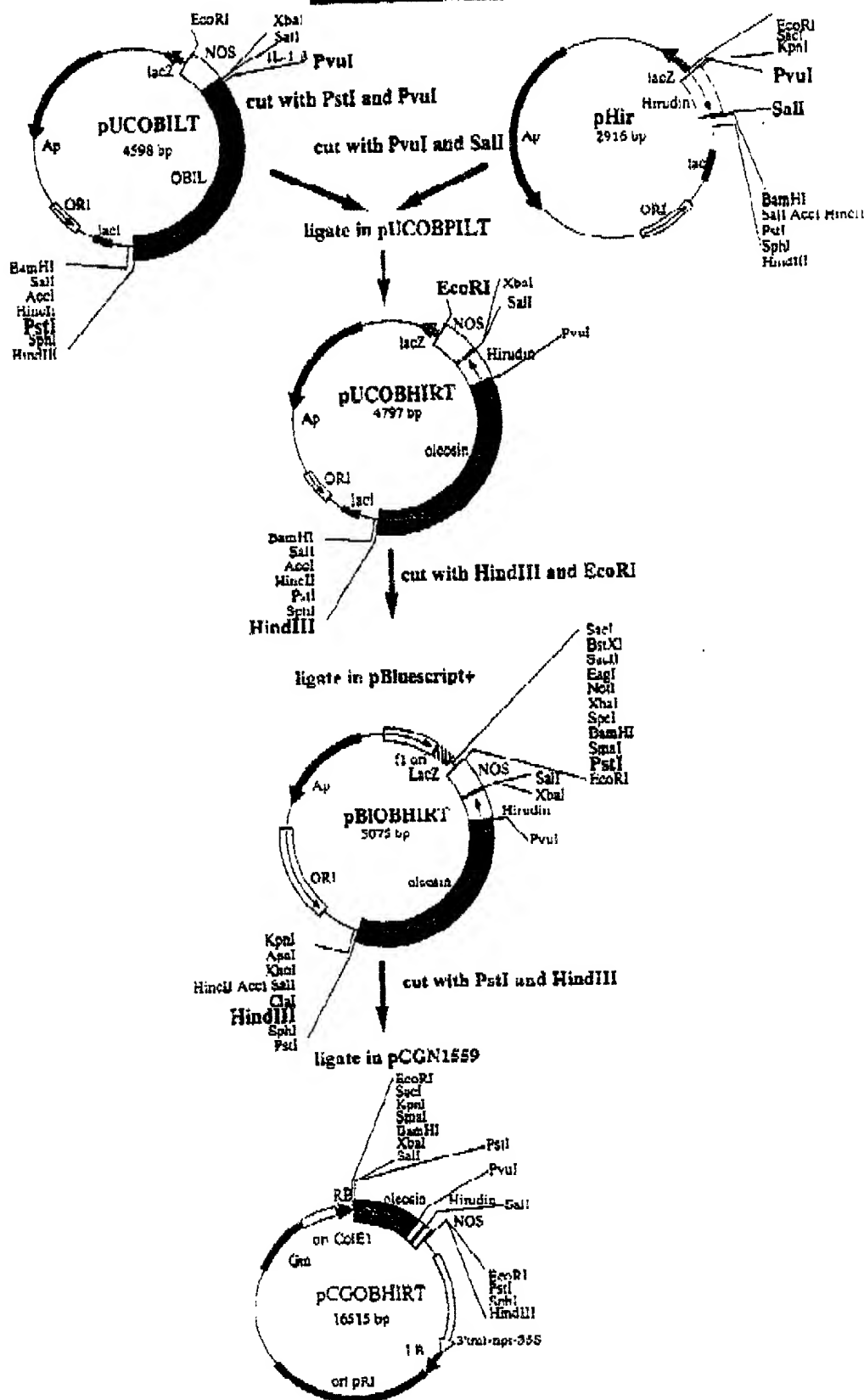
[illegible]

FIGURE 4

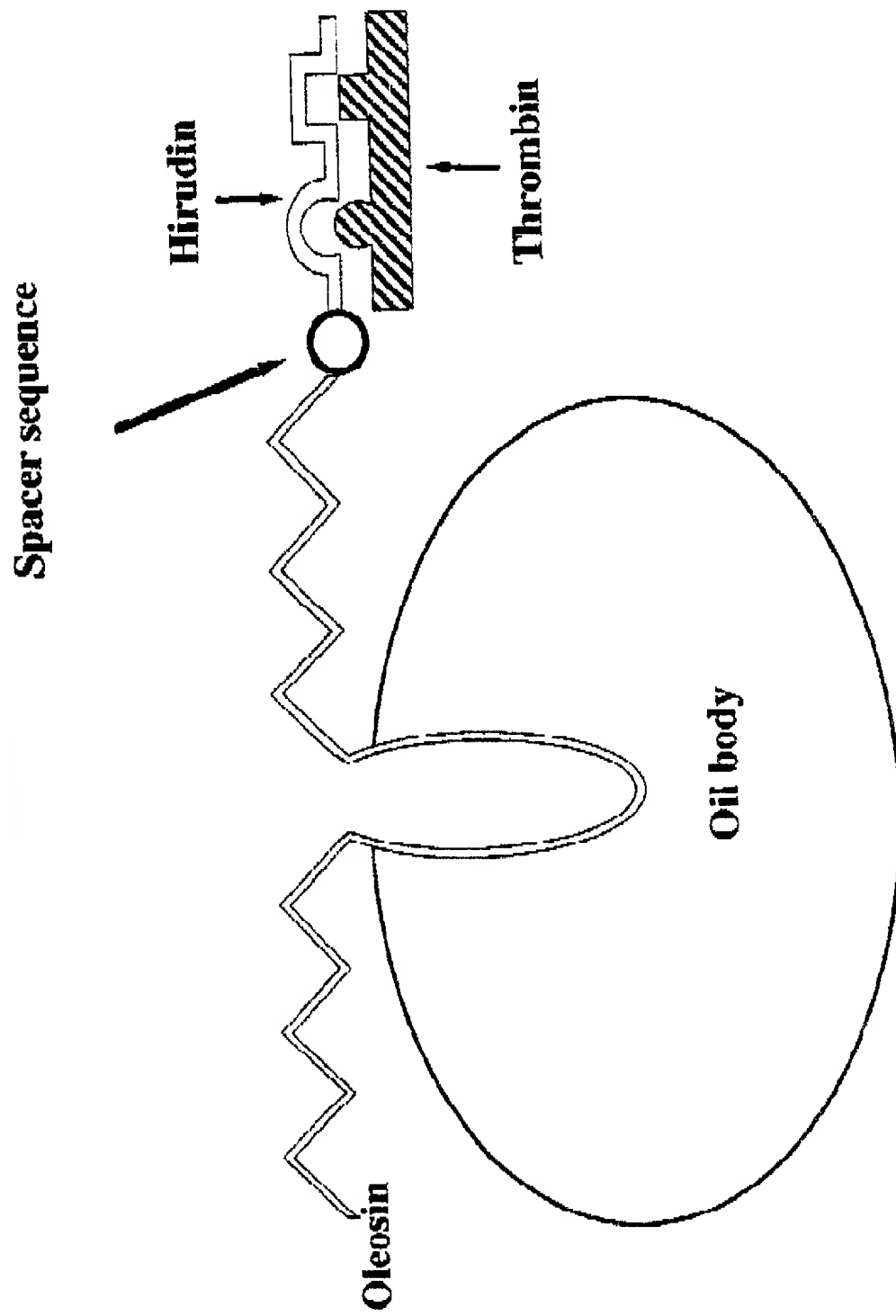


FIGURE 5

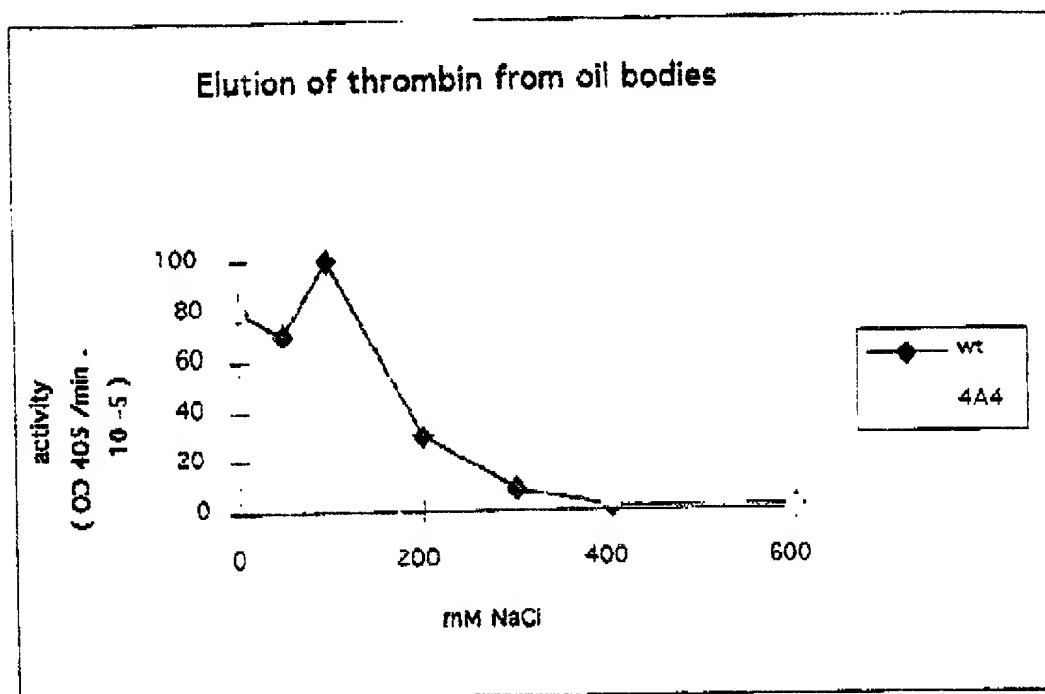


FIGURE 6

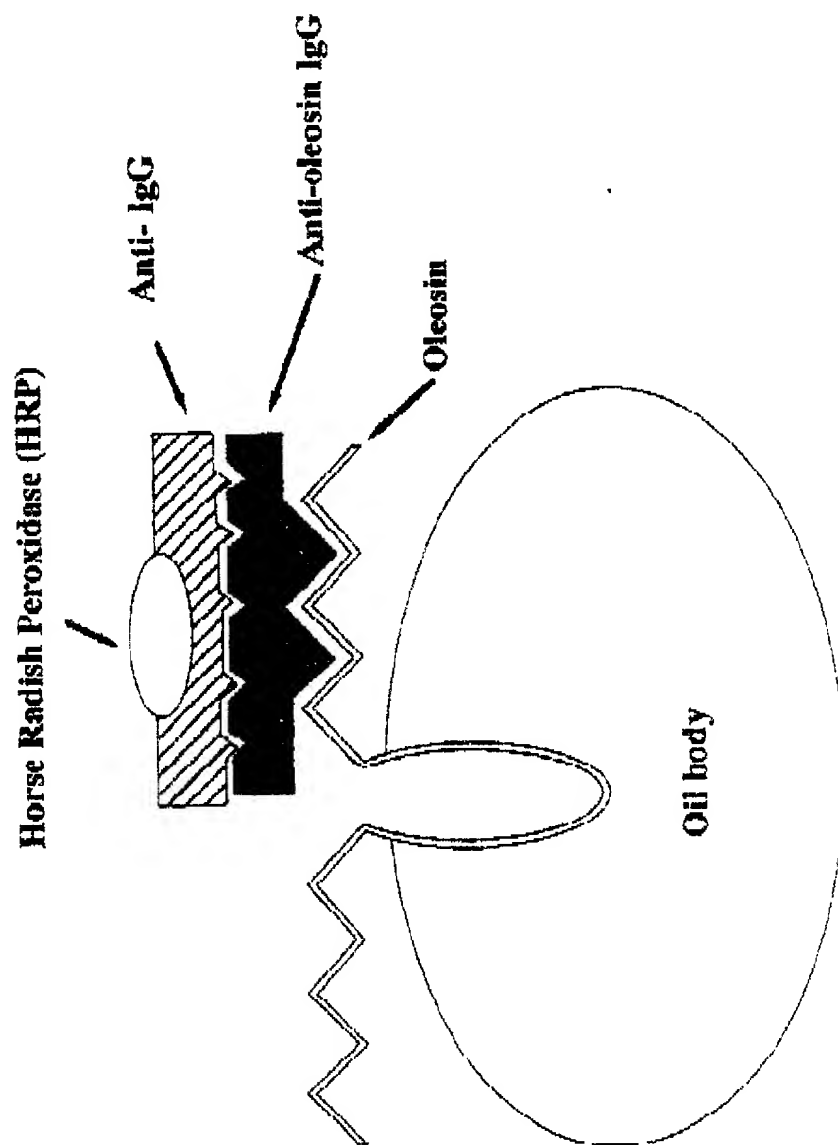


FIGURE 7

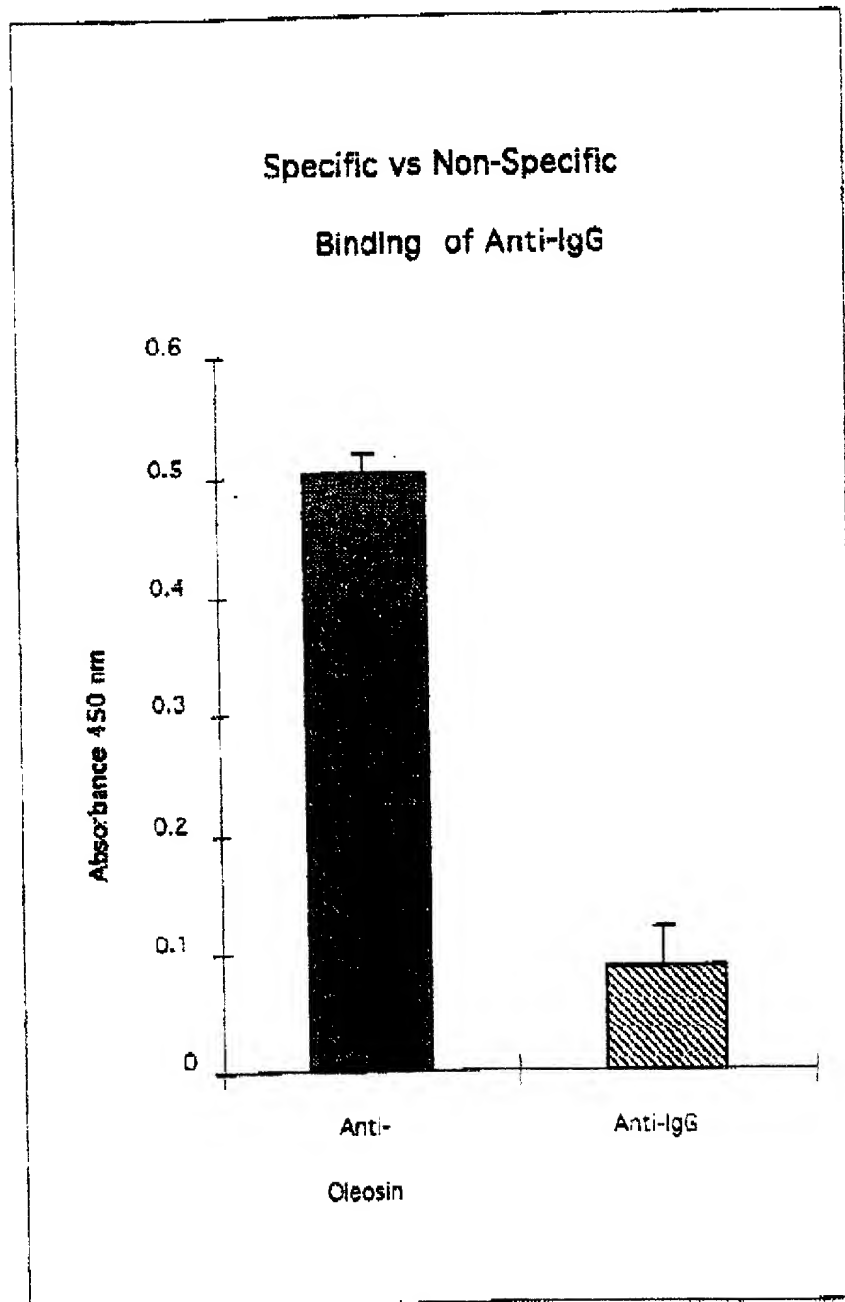


FIGURE 8B

1713 TGC GCC GGC TGC TGC AAG TGC AAA GAG TGC AAA TGC ACC TGC TGC AAG AAA AGC TGC TGC 1772
 208 C A G S C K C K E C K C T S C K K S C C 227

1773 TCC TGC TGT CCT CTG GGC TGT GGC AAG TGT GGC CAG GGC TGC ATC TGC AAA GGC GGC TGC 1832
 228 S C C F V G C A K C A Q C C I C K G A S 247

1833 GAC AAG TGC AGC TGC TGT GGC TGA gggccggcgaacccctgcagaaatgagttccaaagatggtcttctgacgaag 1904
 248 D K C S C C A . 255

1905 ttatgttggctgttttttatggaaattttgtttaagcttgttaattgtggaaagaacgctgctggccttctgctgttttttaattgttg 1964

1965 tgaataaagatgtttcccttggatcaactagctattttttccatttggcttctatggtttctagcaacaaacatttttaaatatg 2064

2065 clycttgatgacatgctgctgctgcttttattatttttaacttacccttccacttccagtttcaaaagtgtgtgcattgactctctgtgt 2144

2145 agtttaagatcgagtgaaagttagatttttgtctatattttatttggggctatttggatgtgttaattggtaaaacatggtcttctga 2224

2225 cagcgtacttttttggcttctggtgttgacgttcttcttttaaaacattatagtagcgtctcttgggtgtgttcttcaattgggttca 2304

2305 acaaaggtgacacttaacttggagatggtgttcttccacttggatatttgaacaaagaatttgggtacc 2366

2366 2365 2364 2363 2362 2361 2360 2359 2358 2357 2356 2355 2354 2353 2352 2351 2350 2349 2348 2347 2346 2345 2344 2343 2342 2341 2340 2339 2338 2337 2336 2335 2334 2333 2332 2331 2330 2329 2328 2327 2326 2325 2324 2323 2322 2321 2320 2319 2318 2317 2316 2315 2314 2313 2312 2311 2310 2309 2308 2307 2306 2305 2304 2303 2302 2301 2300 2299 2298 2297 2296 2295 2294 2293 2292 2291 2290 2289 2288 2287 2286 2285 2284 2283 2282 2281 2280 2279 2278 2277 2276 2275 2274 2273 2272 2271 2270 2269 2268 2267 2266 2265 2264 2263 2262 2261 2260 2259 2258 2257 2256 2255 2254 2253 2252 2251 2250 2249 2248 2247 2246 2245 2244 2243 2242 2241 2240 2239 2238 2237 2236 2235 2234 2233 2232 2231 2230 2229 2228 2227 2226 2225 2224 2223 2222 2221 2220 2219 2218 2217 2216 2215 2214 2213 2212 2211 2210 2209 2208 2207 2206 2205 2204 2203 2202 2201 2200 2199 2198 2197 2196 2195 2194 2193 2192 2191 2190 2189 2188 2187 2186 2185 2184 2183 2182 2181 2180 2179 2178 2177 2176 2175 2174 2173 2172 2171 2170 2169 2168 2167 2166 2165 2164 2163 2162 2161 2160 2159 2158 2157 2156 2155 2154 2153 2152 2151 2150 2149 2148 2147 2146 2145 2144 2143 2142 2141 2140 2139 2138 2137 2136 2135 2134 2133 2132 2131 2130 2129 2128 2127 2126 2125 2124 2123 2122 2121 2120 2119 2118 2117 2116 2115 2114 2113 2112 2111 2110 2109 2108 2107 2106 2105 2104 2103 2102 2101 2100 2099 2098 2097 2096 2095 2094 2093 2092 2091 2090 2089 2088 2087 2086 2085 2084 2083 2082 2081 2080 2079 2078 2077 2076 2075 2074 2073 2072 2071 2070 2069 2068 2067 2066 2065 2064 2063 2062 2061 2060 2059 2058 2057 2056 2055 2054 2053 2052 2051 2050 2049 2048 2047 2046 2045 2044 2043 2042 2041 2040 2039 2038 2037 2036 2035 2034 2033 2032 2031 2030 2029 2028 2027 2026 2025 2024 2023 2022 2021 2020 2019 2018 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986 1985 1984 1983 1982 1981 1980 1979 1978 1977 1976 1975 1974 1973 1972 1971 1970 1969 1968 1967 1966 1965 1964 1963 1962 1961 1960 1959 1958 1957 1956 1955 1954 1953 1952 1951 1950 1949 1948 1947 1946 1945 1944 1943 1942 1941 1940 1939 1938 1937 1936 1935 1934 1933 1932 1931 1930 1929 1928 1927 1926 1925 1924 1923 1922 1921 1920 1919 1918 1917 1916 1915 1914 1913 1912 1911 1910 1909 1908 1907 1906 1905 1904 1903 1902 1901 1900 1899 1898 1897 1896 1895 1894 1893 1892 1891 1890 1889 1888 1887 1886 1885 1884 1883 1882 1881 1880 1879 1878 1877 1876 1875 1874 1873 1872 1871 1870 1869 1868 1867 1866 1865 1864 1863 1862 1861 1860 1859 1858 1857 1856 1855 1854 1853 1852 1851 1850 1849 1848 1847 1846 1845 1844 1843 1842 1841 1840 1839 1838 1837 1836 1835 1834 1833 1832 1831 1830 1829 1828 1827 1826 1825 1824 1823 1822 1821 1820 1819 1818 1817 1816 1815 1814 1813 1812 1811 1810 1809 1808 1807 1806 1805 1804 1803 1802 1801 1800 1799 1798 1797 1796 1795 1794 1793 1792 1791 1790 1789 1788 1787 1786 1785 1784 1783 1782 1781 1780 1779 1778 1777 1776 1775 1774 1773 1772 1771 1770 1769 1768 1767 1766 1765 1764 1763 1762 1761 1760 1759 1758 1757 1756 1755 1754 1753 1752 1751 1750 1749 1748 1747 1746 1745 1744 1743 1742 1741 1740 1739 1738 1737 1736 1735 1734 1733 1732 1731 1730 1729 1728 1727 1726 1725 1724 1723 1722 1721 1720 1719 1718 1717 1716 1715 1714 1713 1712 1711 1710 1709 1708 1707 1706 1705 1704 1703 1702 1701 1700 1699 1698 1697 1696 1695 1694 1693 1692 1691 1690 1689 1688 1687 1686 1685 1684 1683 1682 1681 1680 1679 1678 1677 1676 1675 1674 1673 1672 1671 1670 1669 1668 1667 1666 1665 1664 1663 1662 1661 1660 1659 1658 1657 1656 1655 1654 1653 1652 1651 1650 1649 1648 1647 1646 1645 1644 1643 1642 1641 1640 1639 1638 1637 1636 1635 1634 1633 1632 1631 1630 1629 1628 1627 1626 1625 1624 1623 1622 1621 1620 1619 1618 1617 1616 1615 1614 1613 1612 1611 1610 1609 1608 1607 1606 1605 1604 1603 1602 1601 1600 1599 1598 1597 1596 1595 1594 1593 1592 1591 1590 1589 1588 1587 1586 1585 1584 1583 1582 1581 1580 1579 1578 1577 1576 1575 1574 1573 1572 1571 1570 1569 1568 1567 1566 1565 1564 1563 1562 1561 1560 1559 1558 1557 1556 1555 1554 1553 1552 1551 1550 1549 1548 1547 1546 1545 1544 1543 1542 1541 1540 1539 1538 1537 1536 1535 1534 1533 1532 1531 1530 1529 1528 1527 1526 1525 1524 1523 1522 1521 1520 1519 1518 1517 1516 1515 1514 1513 1512 1511 1510 1509 1508 1507 1506 1505 1504 1503 1502 1501 1500 1499 1498 1497 1496 1495 1494 1493 1492 1491 1490 1489 1488 1487 1486 1485 1484 1483 1482 1481 1480 1479 1478 1477 1476 1475 1474 1473 1472 1471 1470 1469 1468 1467 1466 1465 1464 1463 1462 1461 1460 1459 1458 1457 1456 1455 1454 1453 1452 1451 1450 1449 1448 1447 1446 1445 1444 1443 1442 1441 1440 1439 1438 1437 1436 1435 1434 1433 1432 1431 1430 1429 1428 1427 1426 1425 1424 1423 1422 1421 1420 1419 1418 1417 1416 1415 1414 1413 1412 1411 1410 1409 1408 1407 1406 1405 1404 1403 1402 1401 1400 1399 1398 1397 1396 1395 1394 1393 1392 1391 1390 1389 1388 1387 1386 1385 1384 1383 1382 1381 1380 1379 1378 1377 1376 1375 1374 1373 1372 1371 1370 1369 1368 1367 1366 1365 1364 1363 1362 1361 1360 1359 1358 1357 1356 1355 1354 1353 1352 1351 1350 1349 1348 1347 1346 1345 1344 1343 1342 1341 1340 1339 1338 1337 1336 1335 1334 1333 1332 1331 1330 1329 1328 1327 1326 1325 1324 1323 1322 1321 1320 1319 1318 1317 1316 1315 1314 1313 1312 1311 1310 1309 1308 1307 1306 1305 1304 1303 1302 1301 1300 1299 1298 1297 1296 1295 1294 1293 1292 1291 1290 1289 1288 1287 1286 1285 1284 1283 1282 1281 1280 1279 1278 1277 1276 1275 1274 1273 1272 1271 1270 1269 1268 1267 1266 1265 1264 1263 1262 1261 1260 1259 1258 1257 1256 1255 1254 1253 1252 1251 1250 1249 1248 1247 1246 1245 1244 1243 1242 1241 1240 1239 1238 1237 1236 1235 1234 1233 1232 1231 1230 1229 1228 1227 1226 1225 1224 1223 1222 1221 1220 1219 1218 1217 1216 1215 1214 1213 1212 1211 1210 1209 1208 1207 1206 1205 1204 1203 1202 1201 1200 1199 1198 1197 1196 1195 1194 1193 1192 1191 1190 1189 1188 1187 1186 1185 1184 1183 1182 1181 1180 1179 1178 1177 1176 1175 1174 1173 1172 1171 1170 1169 1168 1167 1166 1165 1164 1163 1162 1161 1160 1159 1158 1157 1156 1155 1154 1153 1152 1151 1150 1149 1148 1147 1146 1145 1144 1143 1142 1141 1140 1139 1138 1137 1136 1135 1134 1133 1132 1131 1130 1129 1128 1127 1126 1125 1124 1123 1122 1121 1120 1119 1118 1117 1116 1115 1114 1113 1112 1111 1110 1109 1108 1107 1106 1105 1104 1103 1102 1101 1100 1099 1098 1097 1096 1095 1094 1093 1092 1091 1090 1089 1088 1087 1086 1085 1084 1083 1082 1081 1080 1079 1078 1077 1076 1075 1074 1073 1072 1071 1070 1069 1068 1067 1066 1065 1064 1063 1062 1061 1060 1059 1058 1057 1056 1055 1054 1053 1052 1051 1050 1049 1048 1047 1046 1045 1044 1043 1042 1041 1040 1039 1038 1037 1036 1035 1034 1033 1032 1031 1030 1029 1028 1027 1026 1025 1024 1023 1022 1021 1020 1019 1018 1017 1016 1015 1014 1013 1012 1011 1010 1009 1008 1007 1006 1005 1004 1003 1002 1001 1000 999 998 997 996 995 994 993 992 991 990 989 988 987 986 985 984 983 982 981 980 979 978 977 976 975 974 973 972 971 970 969 968 967 966 965 964 963 962 961 960 959 958 957 956 955 954 953 952 951 950 949 948 947 946 945 944 943 942 941 940 939 938 937 936 935 934 933 932 931 930 929 928 927 926 925 924 923 922 921 920 919 918 917 916 915 914 913 912 911 910 909 908 907 906 905 904 903 902 901 900 899 898 897 896 895 894 893 892 891 890 889 888 887 886 885 884 883 882 881 880 879 878 877 876 875 874 873 872 871 870 869 868 867 866 865 864 863 862 861 860 859 858 857 856 855 854 853 852 851 850 849 848 847 846 845 844 843 842 841 840 839 838 837 836 835 834 833 832 831 830 829 828 827 826 825 824 823 822 821 820 819 818 817 816 815 814 813 812 811 810 809 808 807 806 805 804 803 802 801 800 799 798 797 796 795 794 793 792 791 790 789 788 787 786 785 784 783 782 781 780 779 778 777 776 775 774 773 772 771 770 769 768 767 766 765 764 763 762 761 760 759 758 757 756 755 754 753 752 751 750 749 748 747 746 745 744 743 742 741 740 739 738 737 736 735 734 733 732 731 730 729 728 727 726 725 724 723 722 721 720 719 718 717 716 715 714 713 712 711 710 709 708 707 706 705 704 703 702 701 700 699 698 697 696 695 694 693 692 691 690 689 688 687 686 685 684 683 682 681 680 679 678 677 676 675 674 673 672 671 670 669 668 667 666 665 664 663 662 661 660 659 658 657 656 655 654 653 652 651 650 649 648 647 646 645 644 643 642 641 640 639 638 637 636 635 634 633 632 631 630 629 628 627 626 625 624 623 622 621 620 619 618 617 616 615 614 613 612 611 610 609 608 607 606 605 604 603 602 601 600 599 598 597 596 595 594 593 592 591 590 589 588 587 586 585 584 583 582 581 580 579 578 577 576 575 574 573 572 571 570 569 568 567 566 565 564 563 562 561 560 559 558 557 556 555 554 553 552 551 550 549 548 547 546 545 544 543 542 541 540 539 538 537 536 535 534 533 532 531 530 529 528 527 526 525 524 523 522 521 520 519 518 517 516 515 514 513 512 511 510 509 508 507 506 505 504 503 502 501 500 499 498 497 496 495 494 493 492 491 490 489 488 487 486 485 484 483 482 481 480 479 478 477 476 475 474 473 472 471 470 469 468 467 466 465 464 463 462 461 460 459 458 457 456 455 454 453 452 451 450 449 448 447 446 445 444 443 442 441 440 439 438 437 436 435 434 433 432 431 430 429 428 427 426 425 424 423 422 421 420 419 418 417 416 415 414 413 412 411 410 409 408 407 406 405 404 403 402 401 400 399 398 397 396 395 394 393 392 391 390 389 388 387 386 385 384 383 382 381 380 379 378 377 376 375 374 373 372 371 370 369 368 367 366 365 364 363 362 361 360 359 358 357 356 355 354 353 352 351 350 349 348 347 346 345 344 343 342 341 340 339 338 337 336 335 334 333 332 331 330 329 328 327 326 325 324 323 322 321 320 319 318 317 316 315 314 313 312 311 310 309 308 307 306 305 304 303 302 301 300 299 298 297 296 295 294 293 292 291 290 289 288 287 286 285 284 283 282 281 280 279 278 277 276 275 274 273 272 271 270 269 268 267 266 265 264 263 262 261 260 259 258 257 256 255 254 253 252 251 250 249 248 247 246 245 244 243 242 241 240 239 238 237 236 235 234 233 232 231 230 229 228 227 226 225 224 223 222 221 220 219 218 217 216 215 214 213 212 211 210 209 208 207 206 205 204 203 202 201 200 199 198 197 196 195 194 193 192 191 190 189 188 187 186 185 184 183 182 181 180 179 178 177 176 175 174 173 172 171 170 169 168 167 166 165 164 163 162 161 160 159 158 157 156 155 154 153 152 151 150 149 148 147 146 145 144 143 142 141 140 139 138 137 136 135 134 133 132 131 130 129 128 127 126 125 124 123 122 121 120 119 118 117 116 115 114 113 112 111 110 109 108 107 106 105 104 103 102 101 100 99 98 97 96 95 94 93 92 91 90 89 88 87 86 85 84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67 66 65 64 63 62 61 60 59 58 57 56 55 54 53 52 51 50 49 48 47 46 45 44 43 42 41 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

FIGURE 10

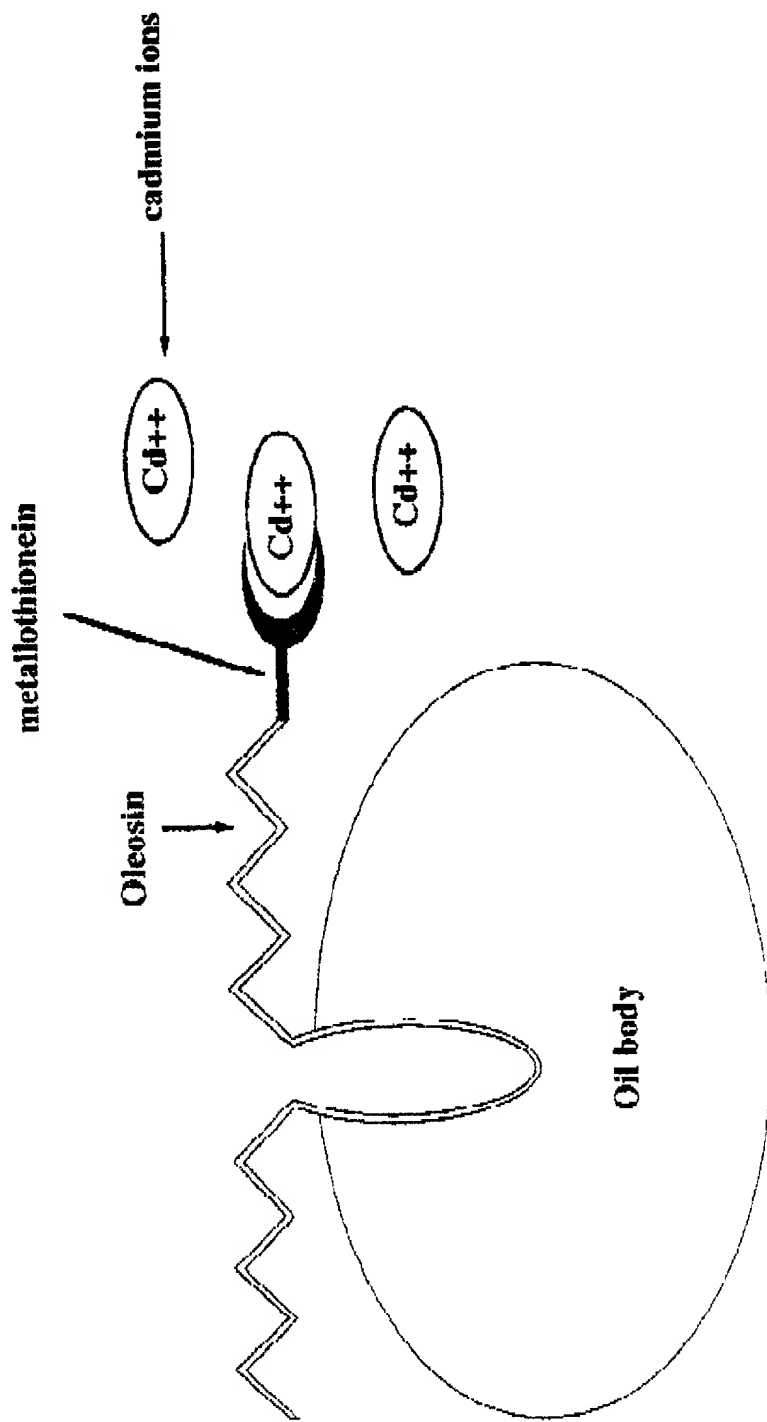
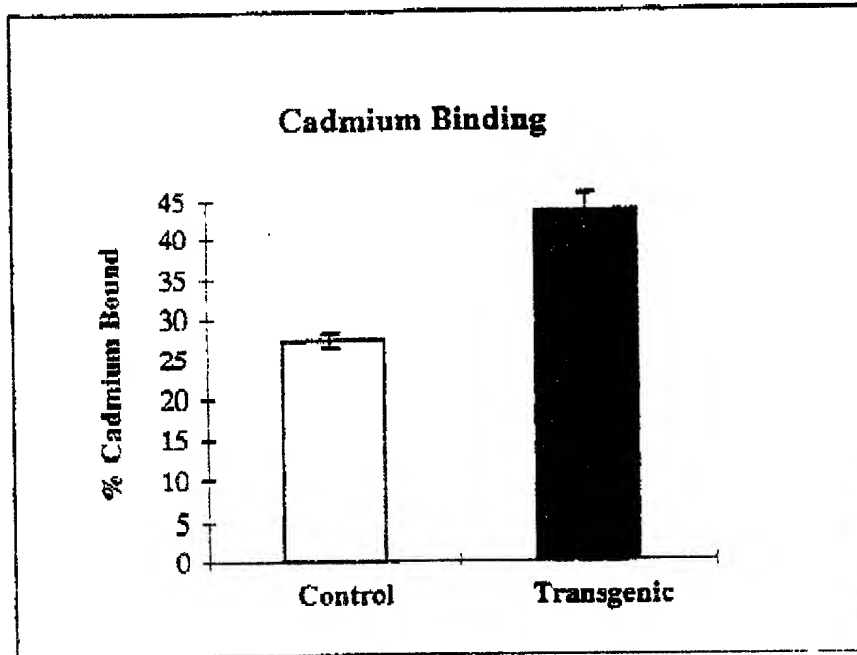


FIGURE 11

A



B

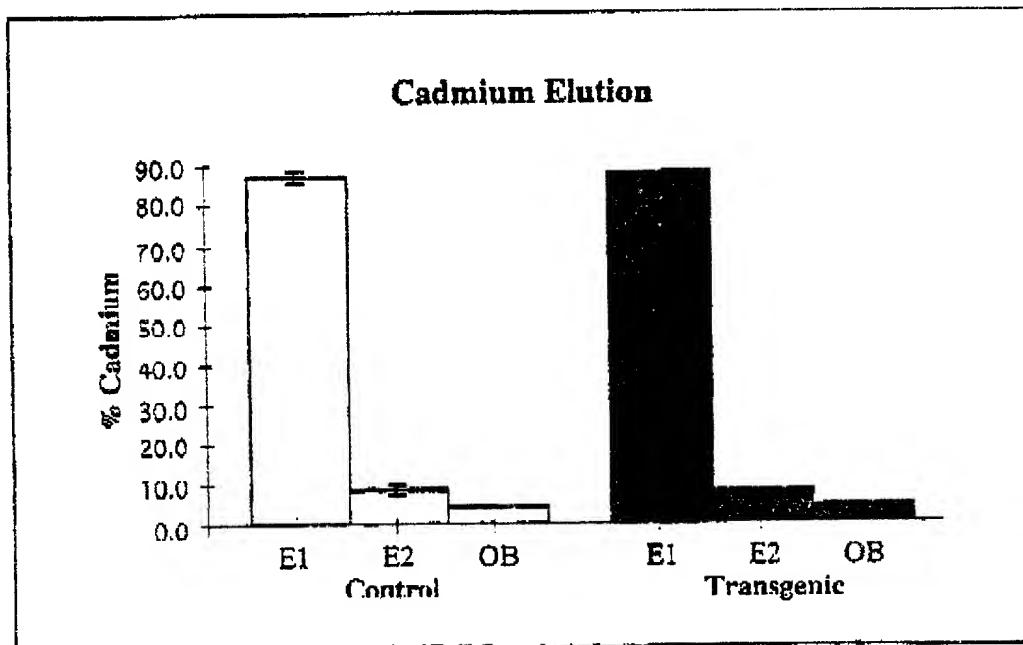
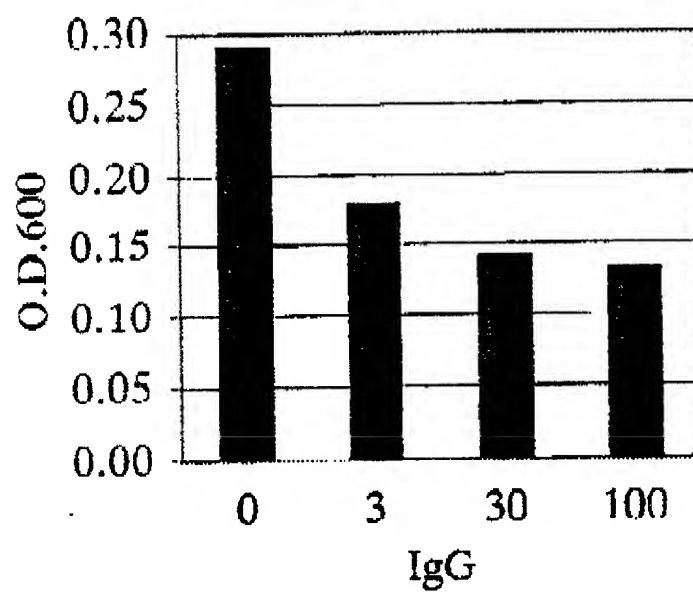


FIGURE 12



[illegible]

804
264

[illegible]

```

1 ccatggctatagcccaacactcggttcggtgcaccaggaaacctccttggttaggcuaagtcccccccgaagaacaacggc 80
      81 gccaaaattgccggaattgctgaacctgagAACGGAACCCATCTTGGGTGCAGCGCAAAATTGCGGTTCAG 160
161 ctgggggttggagagacagacccaalllyagctcgttgaaagggttgtttatgggatcttatatacqqagattggtcgtcga 240
241 gaggcttgagggaacyycaaacatgggtttgggntctggagaaagaaadqtcgggccttagagagagaattgagagytttaga 320
321 yagagatggggogggagatgrrgggadgagaaagcacgaggaacctpgcattatcaaaggagtgacgttgltgaantttggbaac 400
401 ttttaagaygnnagatatagatttcabbatttglatCaatTTTTCTCATTTGTTotagaantgtgcgggaaccaattcllaadacta 480
481 aaactcbaatttttctaatttttgtrgncaatagrtggatAtgtgggcgtatagaagggaalulatbgaggGCCBAAAQCCA 560
561 tactgagcagacdaaagggttocttttycgtttztatgttkoggittcyalyccaaCGEeacattotgagetangrnnnaaac 640
641 aaaogtgottttgantagactctctctgctcauuuatgcageeggtgonatggtgagrnrnttsaacogtggcotacaatt 720
721 gcatsgtgtctccattlgcaulygaststetegtctectfrnkAAtatatcttaaCaAACACTCTtaoctottocaaaaata 800
901 tatacccccttttttgacaaatctonattcBaabaatctcatactctctctatagaacaagAACaanaaa ATG GCG GAT 876
    1                                     H   A   D   J
877 ATA GCT AGA GGA ACC CAT CAC GAT ATC ATC GGC AGA GAC CAG TAC CGG ATT ATG GCC CGA 936
    4 T   A   R   G   T   H   R   D   I   I   G   R   D   Q   Y   P   M   H   G   R   23
937 GAC CGA GAC CAG TAC CAG ATG TCC GGA CGA GSA TCT GAC TAC TCC AAC TCT AGC CAG ATT 996
    24 D   R   D   Q   Y   Q   M   S   G   R   G   S   D   V   S   K   S   P   Q   L   43
997 GCT AAA GCT GCA ACT GCT GTG ACA GCT GGT GCT TCC CTC CTG GTT CTC TCC AGC CTT ACC 1056
    44 A   K   A   A   T   A   V   T   A   C   G   S   T   L   V   L   S   S   L   T   63
1057 CTT GTT GGA ACT CTC ATA GCT TTC ACT GTT GCA ACA CCT CTG CTC GTT ATC TTC AGC CCA 1116
    64 L   V   G   T   V   I   A   L   T   V   A   T   P   L   L   V   I   F   S   P   83
1117 ATC CTT CTC CCG GCT CTC ATC ACA GTT GCA CTC CTC ATC ACC GGT TTT GTT TCC TCT GSA 1176
    84 I   L   V   P   A   L   I   T   V   A   L   L   I   T   G   F   L   S   S   G   103
1177 GGG TTG GGC ATT CCC GCT ATA ACC GTT TTC TTU TGK ATT TAC AA gtanGeasaeatttatcrar 1241
    104 G   F   G   I   A   A   I   T   V   F   S   W   I   Y   X                               118
1242 taettcataattttcgcbaatlolytucatgeatgggttgagcdaghtagnitkgqatcaattxttttgcogaataAcana 1321
1322 lytaaaabaaayaaattgcBAATHCLAgOGaaatatttqtllaacthaatatcqnenttlgacotagetagcttgauvtgt 1401
1402 crngtnatanatocatatataggtGaabaatgcttgytatgatAccatatlgattgtganatag G TAC GCA ADG GGA 1473
    119                                           Y   A   T   U       122
1474 GAG CAC CCA CAG GGA TCA GAC AAG TTG GAC AGT GLA AGE ATG AAG TTG GGA AGC AAA GCT 1533
    123 E   N   P   Q   G   S   D   R   L   D   B   A   R   H   K   L   O   S   K   A   142

```

[illegible]

1534	CAG	GAT	CTG	AAA	GAC	AGA	GCT	CAG	TAC	TAC	GGA	CAG	CAT	ACT	GGT	GGG	GAA	CAT	GAC	1593	
143	Q	D	L	K	D	R	A	Q	Y	Y	G	Q	Q	H	T	G	G	E	H	D	162
1594	CGT	GAC	CGT	ACT	CGT	GGT	GGC	CAG	CAC	ACT	ALT	CTC	GTT	CCA	CGA	GGG	TCC	ATG	GAT	CAA	1653
163	R	D	R	T	R	G	G	Q	H	T	T	L	V	P	R	G	S	H	D	Q	182
1654	CGC	AAT	GCT	TCT	ATC	CAA	AGC	CTT	AAA	CAT	GAT	CCA	AGC	CAA	AGT	GCT	AAC	GTT	TTA	GGT	1713
183	R	N	G	F	X	Q	S	L	K	D	D	P	S	Q	S	A	N	V	L	G	202
1714	GAA	GCT	CAA	AAA	CTT	AAT	GAC	TCT	CAA	GCT	CCA	AAA	GCT	GAT	GGC	CAA	CAA	AAT	AAC	TTC	1773
203	E	A	Q	K	L	N	D	S	Q	A	P	K	A	D	A	Q	Q	N	N	F	222
1774	AAC	AAA	GAT	CAA	CAA	AGC	GCC	TTC	TAT	GAA	ATC	TTG	AAC	ATG	CCT	AAC	TTA	AAC	GAA	GCG	1833
223	N	K	D	Q	Q	S	A	F	Y	E	I	L	N	M	P	N	L	N	E	A	242
1834	CAA	CGT	AAC	GGC	TTC	ATT	CAA	AGT	CTT	AAA	GAC	GAC	CCA	AGC	CAA	AGC	ACT	AAC	GTT	TTA	1893
243	Q	R	N	G	F	I	Q	S	L	K	D	D	P	S	Q	S	T	N	V	L	262
1894	GGT	GAA	GCT	AAA	AAA	TTA	AAC	GAA	TCT	CAA	GCA	CCG	AAA	GCT	GAT	AAC	AAT	TTC	AAC	AAA	1953
263	G	E	A	K	K	L	N	E	S	Q	A	D	K	A	D	N	N	F	N	K	282
1954	GAA	CAA	CAA	AAT	GCT	TTC	TAT	GAA	ATC	TTG	AAT	ATG	CCT	AAC	TTA	AAC	GAA	GAA	CAA	CGC	2013
283	E	Q	Q	N	A	F	Y	E	I	L	N	M	P	N	L	N	E	E	Q	R	302
2014	AAT	GGT	TTC	ATC	CAA	AGC	TTA	AAA	GAT	GAC	CCA	AGC	CAA	AGT	GCT	AAC	CTA	TTG	TCA	GAA	2073
303	N	G	F	I	Q	S	L	K	D	D	P	S	Q	S	A	N	L	L	S	E	322
2074	GCT	AAA	AAG	TTA	AAT	GAA	TCT	CAA	GCA	CCG	AAA	GCG	GAT	AAC	AAA	TTC	AAC	AAA	GAA	CAA	2133
323	A	K	K	L	N	E	S	Q	A	P	K	A	D	N	K	F	N	K	E	Q	342
2134	CAA	AAT	GCT	TTC	TAT	GAA	ATC	TTA	GAT	TTA	GCT	AAC	TTA	AAC	CAA	CAA	CAA	CGC	AAT	GGT	2193
343	Q	N	A	F	Y	E	X	L	H	L	P	N	L	N	E	E	Q	R	N	G	362
2194	TTC	ATC	CAA	AGC	CTA	AAA	GAT	GAT	CCA	AGC	CAA	AGC	GCT	AAC	CTT	TTA	GCA	GAA	GCT	AAA	2253
363	F	I	Q	S	L	K	D	D	P	S	Q	S	A	N	L	L	A	E	A	K	382
2254	AAG	CTA	AAT	GAT	GCT	CAA	GCA	CCA	AAA	GCT	GAC	AAC	AAA	TTC	AAC	AAA	GAA	CAA	CAA	AAT	2313
383	K	L	N	D	A	Q	A	P	K	A	D	N									

FIGURE 15

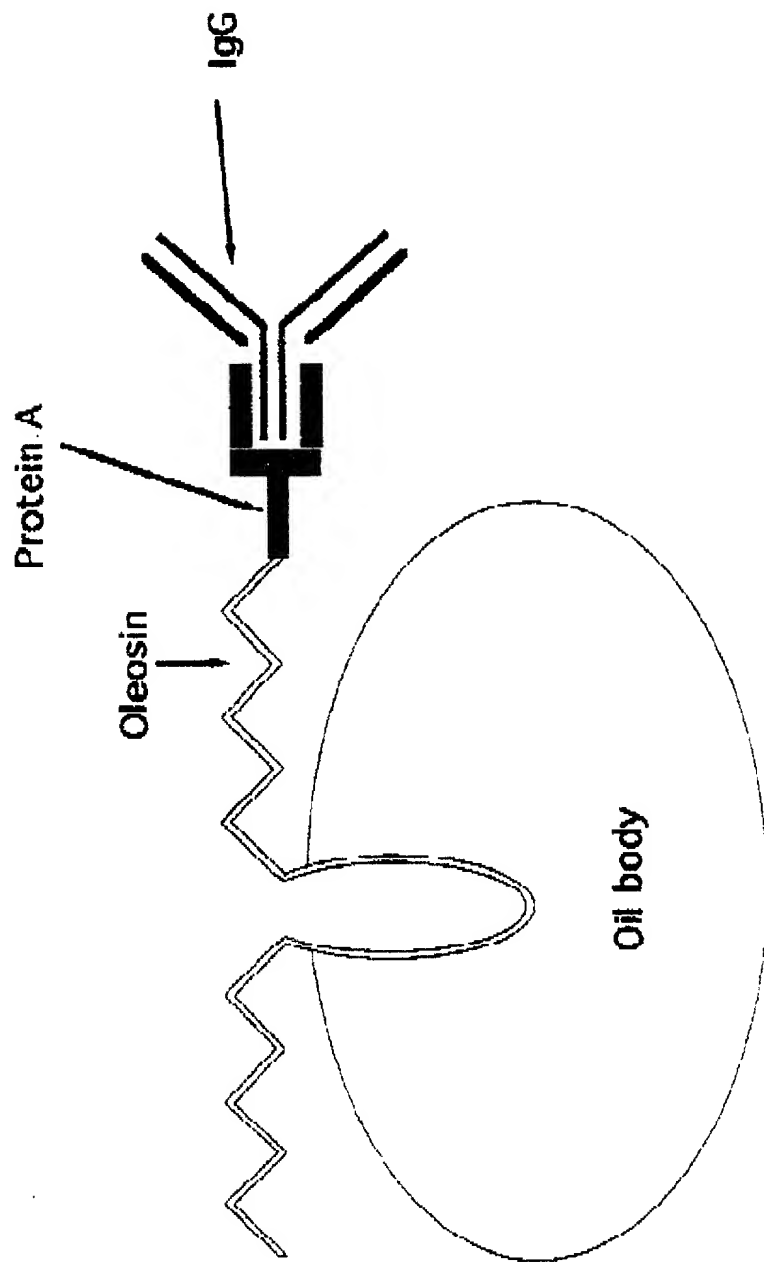


FIGURE 16

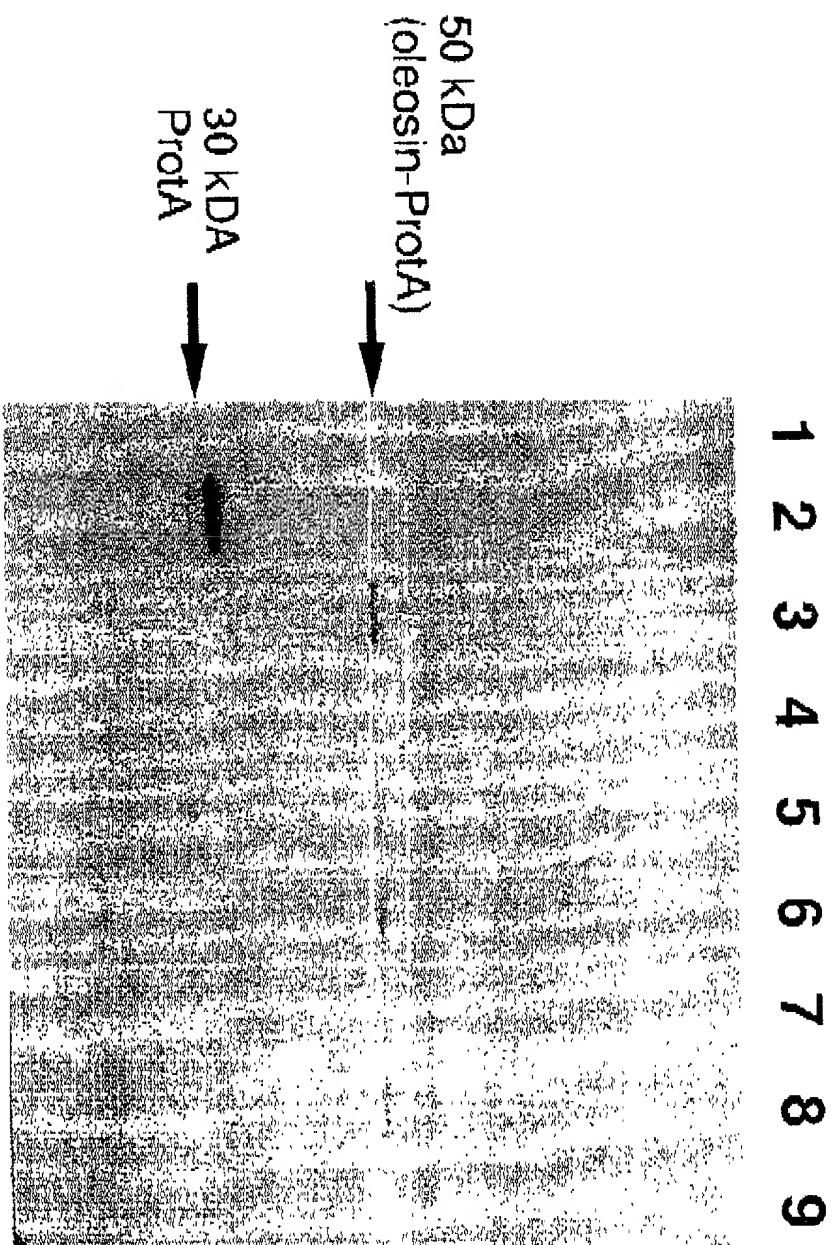
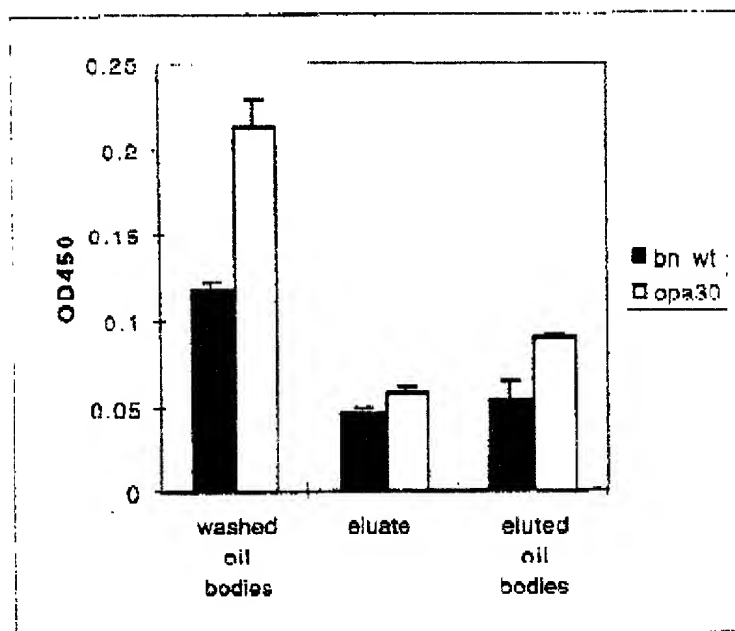


FIGURE 17



[illegible][illegible]

[illegible]

2367	ATG	GCT	GAG	ATC	ATC	CGC	ATT	CCT	CTC	TAC	AAA	SGT	AAG	TCT	CTC	CGT	AAG	GCG	CTG	AAG	2426
277	M	A	F	I	T	R	I	P	L	V	K	G	K	S	L	R	K	A	L	K	291
3427	GAA	CAT	CCA	CTT	CTA	GAA	GAC	TTC	TTC	CAG	AAA	CAA	CAG	TAT	GGC	ATC	ACC	ACC	AAG	TAC	2486
292	E	H	G	L	L	D	D	F	L	Q	X	Q	Q	V	G	T	S	P	K	Y	311
2487	TCC	GGC	TTC	GGT	GAA	GTT	GCT	ACC	GTG	CCA	CTT	ACC	AAC	TAC	CCT	CAT	AGT	CAA	TAC	TTT	2546
312	S	G	F	G	E	V	A	S	V	P	L	T	N	Y	L	D	Q	Q	Y	F	331
2547	GGG	AAG	ATC	TAC	CTC	CGA	ACC	CGS	CCT	CAA	GAG	TTC	ACC	GTT	CTC	CTT	GAT	ACT	GGT	TCC	2606
332	G	K	I	Y	L	G	T	P	P	Q	E	F	T	V	L	F	D	T	G	S	351
2607	TCT	GAC	TTC	TGG	GTT	CCC	TCT	ATC	TAC	TGC	AAC	AGC	AAT	GCC	TGC	AAG	AAC	CAC	CAA	AGA	2666
352	Q	D	F	W	V	P	S	I	Y	C	K	S	N	A	C	K	N	H	Q	R	371
2667	TTC	GAT	CCC	ACA	AAH	TGC	TCC	ACC	TTC	CAG	AAC	TTA	GGC	AAA	CCC	TTG	TCT	ATA	CAC	TAC	2726
372	P	D	P	R	K	S	S	T	F	Q	N	I	G	K	P	L	S	I	H	Y	391
2727	GGT	ACA	GGT	AGC	ATG	CAA	GGA	ATC	TTA	CCC	TAT	GAT	ACC	CTC	ACT	GTG	TCC	AAC	ATT	CTC	2786
392	G	T	G	S	M	Q	S	I	L	G	Y	D	T	V	T	V	S	N	I	V	411
2787	GAC	ATT	CAA	CAG	ACA	GTA	GGA	CTT	AGC	ACC	CAA	AAA	CTA	GGT	GAT	GTG	TTC	ACC	TAT	GCA	2846
412	D	I	O	Q	T	V	G	L	S	T	Q	E	P	G	D	V	F	I	Y	A	431
2847	GAA	TTC	GAT	CCC	ATC	CTT	GGT	ATG	GCA	TAC	CCA	TGG	CTC	GCG	TCA	GAG	TAC	TGC	ATA	CTT	2906
432	E	F	D	G	I	L	G	M	A	Y	P	S	L	A	S	E	Y	S	I	P	451
2907	GTG	TTT	GAC	AAC	ATG	ATG	AAC	CGA	CAC	CTA	GTA	GCT	CAA	GAC	TTG	TTC	CCG	GTT	TAC	ATG	2966
452	V	F	D	N	M	M	N	R	H	I	V	A	Q	D	L	F	S	V	Y	M	471
2967	GAC	AGG	AAT	GGC	CAG	GAG	AGC	ATC	CTC	ACC	CTT	GGA	GCT	ATT	GAT	CCA	TCT	TAC	TAC	ACA	3026
472	D	R	N	Q	Q	E	S	M	L	T	L	G	A	I	D	D	S	Y	Y	T	491
3027	GGA	TCT	CTT	CAC	TGG	GTT	CCA	CTT	ACT	GTG	CAG	CAG	TAC	TGG	CAA	TTC	ACT	GTG	GAC	ACT	3086
492	G	S	L	H	W	V	P	V	T	V	Q	Q	Y	W	Q	F	T	V	D	S	511
3087	GTG	ACC	ATC	AGC	GCT	GTG	GTT	CTT	GCA	TGT	GAA	GGT	GGA	TGT	CAA	CCT	ATC	TTG	GAT	ACC	5146
512	V	T	I	S	G	V	V	V	A	C	E	G	G	C	Q	A	I	L	D	T	531
3147	GGT	ACC	TCC	AAG	CTG	GTG	GCA	CCT	AGC	AGC	GAC	ATT	CTC	AAC	ATT	CAG	CAA	GCT	ATT	GGA	3206
532	C	T	S	K	I	V	C	P	S	S	D	I	L	N	I	Q	Q	A	I	G	551
3207	GCC	ACA	CAG	AAC	CAC	TAC	CCT	GAG	TTG	GAC	ATA	GAT	TGC	CAC	AAC	CTT	AGC	TAC	ATG	CCT	3266
552	A	T	Q	N	Q	Y	G	F	F	D	I	D	C	D	N	L					

FIGURE 18C

```
4080 aatctctatttctctctctcagataaaaaaaataattacagagttgggttgataaaatattcgaggatctttaaataataataa 4159
4160 taataaaataaataaataatctatctatataaacttaattataataataaactttatctataaaaaagttaaacattgtcatata 4239
4240 tctatctatctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 4319
4320 acatactctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 4399
4400 gtgtctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 4479
4480 aatttgggttctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 4559
4560 accgtgtctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 4639
4640 tctatctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 4698
```

Figure 18C shows a sequence of nucleotides (A, T, C, G) arranged in a grid-like pattern. The sequence is presented in a single line of text, with line numbers 4080, 4160, 4240, 4320, 4400, 4480, 4560, and 4640 indicating the start of each line. The sequence ends with a line number 4698. The sequence is a continuous string of nucleotides, with spaces used to separate the lines for readability. The sequence is a single line of text, with line numbers 4080, 4160, 4240, 4320, 4400, 4480, 4560, and 4640 indicating the start of each line. The sequence ends with a line number 4698.